



UNIVERSIDADE FEDERAL DO MARANHÃO

Programa de Pós-Graduação em Ciência da Computação

Robert Douglas de Araujo Santos

Uma abordagem para diagnóstico automático de estrabismo baseado em vídeos do exame Cover test alternado utilizando Deep Learning

**São Luís
2022**

Robert Douglas de Araujo Santos

**Uma abordagem para diagnóstico automático de
estrabismo baseado em vídeos do exame *Cover test*
alternado utilizando *Deep Learning***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFMA, como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Maranhão - UFMA

Centro de Ciências Exatas e Tecnologia - CCET

Programa de Pós-Graduação em Ciência da Computação - PPGCC

Orientador: Prof. Dr. João Dallyson Sousa de Almeida

São Luís - MA

2022

Robert Douglas de Araujo Santos

**Uma abordagem para diagnóstico automático de
estrabismo baseado em vídeos do exame *Cover test*
alternado utilizando *Deep Learning***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFMA, como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Trabalho aprovado em São Luís - MA, 23 de Junho de 2022:

**Prof. Dr. João Dallyson Sousa de
Almeida**
Orientador

Prof. Dr. Geraldo Braz Júnior
(Avaliador Interno)

Prof. Dr. Sidnei Alves de Araújo
(Avaliador Externo)

**Prof. Dr. Jorge Antônio Meireles
Teixeira**
(Avaliador Externo)

São Luís - MA
2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Santos, Robert Douglas de Araujo.

Uma abordagem para diagnóstico automático de estrabismo baseado em vídeos do exame Cover test alternado utilizando Deep Learning / Robert Douglas de Araujo Santos. - 2022.
75 p.

Orientador(a): João Dallyson Sousa de Almeida.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, São Luís, 2022.

1. Covert test alternado. 2. Diagnóstico automático.
3. Estrabismo. 4. YOLOv5. I. Almeida, João Dallyson Sousa de. II. Título.

A família e amigos

Agradecimentos

Agradeço ao meus pais, Irismar e Nonato, por sempre me darem suporte e incentivo para continuar estudando apesar das dificuldades encontrada e realizar mais essa etapa. E a minha namorada Letícia por todo companheirismo, incentivo e ajuda no decorrer dessa jornada.

Agradeço ao meu orientador, João Dallyson Sousa de Almeida, que além dos vários ensinamentos foi sempre muito compreensivo com as minhas dificuldades e ajudou no que fosse possível.

Ao NCA, laboratório que me deu todo suporte necessário e possível para desenvolver esta pesquisa. Ao PPGCC-UFMA e todo seu corpo docente, por transmitir conhecimentos importantíssimos para a minha evolução acadêmica e profissional.

À todos os meus amigos, que me apoiaram em momentos difíceis e desesperançosos da vida pessoal ao meio de uma pandemia, vocês fizeram que essa o término dessa etapa fosse possível. Em especial aos grandes amigos Victor (Vitu), Ítalo e Pedro (Pedin).

A todos os integrantes da equipe do projeto PPE2 do instituto Tecgraf pela confiança, receptividade, oportunidade e compreensão. Esse apoio me permitiu conciliar o início da minha carreira profissional com a etapa final do mestrado.

Por fim, agradeço a todas as pessoas que passaram durante essa jornada e ajudaram a moldar a pessoa e profissional que sou hoje.

*"Time for a new master."
(Lifestealer, Defence of the Ancients 2)*

Resumo

O estrabismo é um distúrbio em que a linha de visão de um dos olhos não consegue fixar no objeto alvo que os dois olhos estão visualizando, ou seja, enquanto um dos olhos fixa o olhar no objeto, a linha de visão do outro está para uma direção diferente. O estrabismo afeta uma parcela crescente da população adulta e grande parte das crianças. Pode causar problemas estéticos e perda de visão, que em muitos casos pode ser prevenida. Considerando as características do distúrbio, o número de médicos especializados disponíveis para a população e o avanço computacional nos últimos anos, este trabalho apresenta um método computacional para diagnosticar estrabismo usando uma rede neural convolucional. Para isso utiliza a YOLOv5 para detecção dos olhos e ocluser em vídeos digitais do exame *Cover Test* alternado. O método foi avaliado em duas bases de vídeos, a primeira intitulada de Dataset CV-I que possui 13 vídeos de voluntários adquiridos em consultório médico no hospital da Universidade Federal do Maranhão. A segunda base chamada de Dataset CV-II, composta por vídeos de 57 voluntários e que foram adquiridos em escolas públicas na cidade de São Luís, Maranhão. Os resultados obtidos pelo método proposto são comparados com medições de um médico especialista. De acordo com os resultados, o método proposto detecta corretamente os olhos 100% das vezes e o ocluser 97% das vezes no conjunto de dados de treinamento e teste. A assertividade da rede neural na detecção dos olhos também foi validada utilizando uma base de vídeos de paralisia do sexto nervo abreviada como PSN, essa base conta com 35 vídeos adquiridos em ambientes variados. Nessa base totaliza-se 19.966 olhos, dos quais a rede identificou corretamente 19.488, errando apenas 478 e com uma assertividade total de 97,61%. O método também obteve um erro máximo de 7,28 dioptrias prismáticas(Δ) na aferição de desvios horizontais, ficando abaixo do limiar de erro definido na literatura como 8Δ . Além disso, ao se avaliar o resultado no Dataset CV-I o método alcança uma acurácia igual a 100% e (95%IC = [0,77; 1]), 100% de especificidade com (95%CI = [0,66; 1]) e 100% de sensibilidade com (95%IC = [0,48; 1]) para o estrabismo horizontal. Para o Dataset CV-II o método obteve uma sensibilidade de 66,67% com (IC 95% [0,64; 0,99]), especificidade de 100% com (IC 95% [0,93; 1]) e acurácia de 99,33% com (IC 95% [0,92; 1]) quando considerada apenas a medição horizontal. Por fim, ainda para o Dataset CV-II, ao avaliar as medições horizontais e verticais em conjunto, o método obteve uma sensibilidade de 66,67% com (IC 95% [0,63; 0,99]), especificidade de 94,44% com (IC 95% [0,84; 0,98]) e acurácia de 93,89% com (IC 95% [0,84; 0,98]).

Palavras-chave: Estrabismo, Diagnóstico automático, *Covert test* alternado, YOLOv5.

Abstract

A condition known as strabismus occurs when one eye's line of sight is unable to fix on the target item that both eyes are picturing; rather, when one eye fixes its gaze on the object, the other eye's line of sight is in a different direction. Strabismus affects a growing portion of the adult population and a large proportion of children. It can cause aesthetic issues and vision loss, which can be avoided in many cases. Taking into account the characteristics of the disturb, the number of specialized doctors available to the public, and computer advancements in recent years, this study presents a computational method for detecting and diagnosing strabismus using a neural convolutional network YOLOv5 for ocular detection and occluder in digital exam videos of *Cover Test*. The method was carried out on two video databases, the first of which is called Dataset CV-I and contains 13 volunteer videos obtained in the hospital the Universidade Federal do Maranhão. The second basis, known as Dataset CV-II, is made up of 57 volunteer videos that were acquired in public schools in the city of São Luís, Maranhão. The results obtained by the proposed method are purchased with studies by a specialist doctor. According to the results, the proposed method correctly detects the eyes 100% of the time and correctly occluder 97% of the time in the training and testing dataset. Using the video base of sixth nerve palsy, also known as PSN, which comprises 35 movies collected in various contexts, the assertiveness of the neural network in identifying the eyes was also evaluated. accordingly, there are 19,966 eyes in total, of which the network correctly detected 19,488, with just 478 errors, and a total assertiveness of 97.61%. In the measurement of horizontal deviations, the method yielded a maximum error of 7.28Δ , falling short of the literature-defined error limit of 8Δ . In addition, when evaluating the result in the CV-I Dataset, the method achieves an accuracy equal to 100% e (95%CI = [0.77; 1]), 100% de specificity with (95%CI = [0.66; 1]) and 100% of sensitivity with (95%CI = [0.48; 1]) for horizontal strabismus. For the CV-II Dataset, the method obtained a sensitivity of 66.67% with (CI 95% [0.64; 0.99]), specificity of 100% with (CI 95% [0.93; 1]) and accuracy of 99.33% with (CI 95% [0.92; 1]) when considering only the horizontal measurement. Finally, when considering the horizontal and vertical measurements together, the method obtained a sensitivity of 66.67% with (CI 95% [0.63; 0.99]), specificity of 94.44% with (CI 95% [0.84; 0.98]) and accuracy of 93.89% with (CI 95% [0.84; 0.98]).

Keywords: Strabismus, Automatic diagnosis, Alternate *Covert Test*, YOLOv5.

Lista de ilustrações

Figura 1 – Estrutura detalhada do olho humano.	29
Figura 2 – Estruturas visíveis do olho.	29
Figura 3 – Exemplificação da visão binocular.	31
Figura 4 – Estrabismo conforme direção do desvio. (A) Exotropia (B) Esotropia (C) Hipertropia (D) Hipotropia.	32
Figura 5 – Execução do <i>Cover test</i>	33
Figura 6 – Execução do <i>Cover test</i> alternado.	33
Figura 7 – Etapas da arquitetura da <i>perceptron</i>	34
Figura 8 – Arquitetura básica <i>Multilayer perceptron</i>	35
Figura 9 – Arquitetura base de uma CNN.	38
Figura 10 – Execução de uma convolução.	39
Figura 11 – Execução do <i>Max pooling</i>	39
Figura 12 – <i>Pipeline de execução da YOLO</i>	41
Figura 13 – Arquitetura da YOLOv5.	42
Figura 14 – Diagrama do <i>backbone</i> YOLOv5.	44
Figura 15 – Camada de foco.	44
Figura 16 – Bloco residual parcial da CSP Net	45
Figura 17 – Estrutura do bloco CBL.	45
Figura 18 – Bloco CSP1_X.	46
Figura 19 – Estrutura da <i>Spatial Pyramid Pooling</i>	46
Figura 20 – Estrutura do <i>Neck</i> da YOLOv5.	47
Figura 21 – Bloco CSP2_X.	47
Figura 22 – Visão geral do método proposto.	52
Figura 23 – Esquema de marcação das imagens.	56
Figura 24 – Detecção da região do limbo de ambos os olhos.	57
Figura 25 – Definição do centro do limbo.	57
Figura 26 – Processo de oclusão de um olho.	59
Figura 27 – Comparação entre medidas de especialistas, Valente et al. (2017) e o método proposto (MDH).	63
Figura 28 – Número de pacientes por intervalo de erro para a medida horizontal (MDH) e para a medida vertical (MDV).	65

Lista de tabelas

Tabela 1 – Trabalhos selecionados após 3 ^a fase.	21
Tabela 2 – Avaliação de qualidade do trabalhos	22
Tabela 3 – Resumo dos trabalho relacionados	27
Tabela 4 – Resultado modelo YOLO no <i>Dataset</i> PSN	61
Tabela 5 – Comparação entre medidas do especialista e medição realizada pelo método proposto	62
Tabela 6 – Comparação entre o diagnósticos de pacientes doentes.	64
Tabela 7 – Comparação entre a avaliação do especialista e os resultados alcançados.	67
Tabela 8 – Produções científicas desenvolvidas durante a pesquisa.	71

Lista de abreviaturas e siglas

AC - Acurácia

CNN - *Convolutional Neural Network*

CSP - *Cross Stage Partial*

CTA - *Cover test* alternado

FN - Falso Negativo

FP - Falso Positivo

FPS - *Frames per second*

GIoU - *Generalized Intersection over Union*.

IAPB - Agência para prevenção de Cegueira

IoU- *Intersection over Union*

MAE - *Mean Absolute Error*

MLP - *Multilayer Perceptron*

OMS - Organização Mundial de Saúde

R-FCN - *Region-based Fully Convolutional Networks*

RNA - Redes neurais artificiais

RSL - Revisão Sistemática da Literatura

SVM - *Support Vector Machine*

UFMA - Universidade Federal do Maranhão

VN - Verdadeiro Negativo

VP - Verdadeiro Positivo

YOLO - *You only look once*

YOLOv5 - *You only look once version 5*

Sumário

1	INTRODUÇÃO	15
1.1	Objetivo Geral	17
1.2	Objetivos específicos	17
1.3	Organização da Dissertação	17
2	TRABALHOS RELACIONADOS	19
2.1	Revisão Sistemática da Literatura	19
2.1.1	Resultados da RSL	20
2.2	Detecção de estrabismo	22
2.3	Diagnóstico de Estrabismo	24
2.4	Considerações	26
3	FUNDAMENTAÇÃO TEÓRICA	28
3.1	O Olho humano	28
3.1.1	Visão Binocular	30
3.2	O Estrabismo	30
3.3	Cover Test	32
3.4	Redes neurais artificiais	33
3.4.1	<i>Multilayer Perceptron</i>	35
3.4.2	<i>Backpropagation</i>	35
3.4.3	Redes neurais convolucionais	37
3.4.4	YOLO - <i>You Only Look Once</i>	40
3.4.5	YOLO Versão 5	42
3.4.5.1	Camada de entrada	43
3.4.5.2	<i>Backbone</i>	44
3.4.5.3	Estrutura CSP - <i>CSP Network</i>	45
3.4.5.4	<i>Neck</i> - Pescoço	47
3.4.5.5	Camada de saída	48
3.5	Métricas de avaliação	48
3.6	Testes de hipóteses	50
3.7	Considerações	51
4	MATERIAIS E MÉTODO	52
4.1	Bases de vídeos	52
4.1.1	Dataset de videos cover test I (CV-I)	52
4.1.2	Dataset de videos cover test II (CV-II)	53

4.1.3	Dataset de vídeos paralisia sexto nervo (PSN)	54
4.2	Localização e rastreamento dos olhos e ocluser	54
4.2.1	Treinamento da rede YOLOv5	55
4.2.2	Deteção do centro dos olhos	56
4.3	Cálculo do desvio	58
4.4	Hardware e Software utilizados	59
5	RESULTADOS E DISCUSSÃO	61
5.1	Resultados dos detecção dos olhos	61
5.2	Resultados do diagnóstico para o Dataset CV-I	61
5.3	Resultados do diagnóstico para o Dataset CV-II	63
5.4	Discussão	68
6	CONCLUSÃO	69
6.1	Trabalhos futuros	70
6.2	Contribuições	70
6.3	Produções científicas	71
	REFERÊNCIAS	72

1 INTRODUÇÃO

Um levantamento realizado pela Organização Mundial de Saúde (OMS) na década de 90 apurou que existiam cerca de 38 milhões de pessoas cegas, 110 milhões com baixa acuidade visual, aproximadamente 7 milhões de pessoas perdiam a visão por ano e que esse valor aumentava entre 1 e 2 milhões a cada ano ([ORGANIZATION et al., 2000](#)). Além disso, o mesmo levantamento chegou a conclusão que pelo menos dois terços dos casos de cegueira são evitáveis ou tratáveis. Considerando esse cenário, a OMS e a Agência para prevenção de Cegueira lançaram um plano conjunto intitulado de “VISÃO 2020: O direito a Visão”. Essa iniciativa pretendia eliminar a cegueira evitável ou tratável até o ano de 2020 adotando como estratégia a assistência oftalmológica básica, que por sua vez integrava três fatores: o desenvolvimento de infra-estrutura e tecnologias, controle específico de doenças e desenvolvimento de recursos humanos. No entanto, alguns fatores dificultaram que o plano alcançasse o objetivo proposto, dentre eles, pode-se destacar a relação entre o desenvolvimento de recursos humanos, oftalmologistas por exemplo, e o aumento da população, que em 2009 era em torno de 6.5 bilhões e que até 2050 estima-se alcançar 9 bilhões de pessoas ([BOARETTO, 2009](#)).

Em 2019 outro relatório da OMS sobre a visão reforçou o alerta acerca do problema de cegueira evitável, principalmente em crianças. Nesse relatório foi mostrado que havia 1,4 milhão de crianças cegas no mundo. A prevalência do estrabismo em crianças variava de 0.3/1000 crianças de 0-15 anos em países desenvolvidos para 1.5/1000 crianças em países muito pobres. O mesmo relatório mostrou que 500.000 crianças ficam cegas por ano (quase uma por minuto) ([WHO, 2019](#)). Outro estudo anterior ao relato da OMS, corrobora que boa parte das crianças cegas do mundo são cegas devido a causas evitáveis, sendo 15% tratáveis e 28% preveníveis ([ÁVILA; ALVES; NISHI, 2015](#)).

No ano de 2021, um outra pesquisa avaliou novamente as tendências de prevalência de cegueira e deficiência da visão nos últimos 30 anos. Esse estudo foi uma iniciativa de contribuir com a OMS e seu plano já mencionado. Foi constatado que em 2020, aproximadamente 43,3 milhões de pessoas eram cegas, 295 milhões de pessoas tinham alguma deficiência visual e que o número de pessoas cegas aumentou em 50,6% ([BOURNE et al., 2021](#)). Esse mesmo estudo ainda realizou uma estimativa de que em 2050 por volta de 61 milhões de pessoas serão cegas, 474 milhões terão deficiência visual moderada e grave e 360 milhões terão deficiência visual leve.

Dentre as causas de cegueira que levam ao comprometimento sensorial da visão pelo cérebro está o estrabismo. Trata-se de uma condição em que a linha de visão de um dos olhos não consegue fixar no objeto alvo que os dois olhos estão visualizando, ou seja,

enquanto um dos olhos fixa o olhar no objeto, a linha de visão do outro está para uma direção diferente. O estrabismo também pode ser descrito considerando duas condições de manifestação. No estrabismo manifesto, também conhecido como tropia, o desvio da linha de visão dos olhos pode ser observado apenas olhando para o paciente. Por sua vez, o latente - também dito foria - diferente do anterior, só pode ser observado quando a linha de visão dos olhos é interrompida. Para ambos, exames como o teste de Hirschberg ¹, o teste de tela de Hess Lancaster² e o Cover test são necessários para o diagnóstico. No entanto, pode ser prevenido com o exame oftalmológico das crianças, nos primeiros anos de vida, com até aproximadamente três anos de idade.

Considerando o panorama apresentado, em que a quantidade de médicos especializados em oftalmologia tem dificuldades em atender a alta demanda populacional, que segundo levantamento do Conselho brasileiro de Oftalmologia (CBO), para uma população brasileira de 211.755.692 milhões de habitantes existiam em 2021, apenas 19.471 indivíduos médicos oftalmologistas (CBO, 2021). Ainda nesse censo realizado pelo CBO, dentre esses profissionais 1.798 indivíduos médicos atuam em mais de uma região, 3.250 atuam em mais de um estado e, 8.748 atuam em mais de um município (CBO, 2021). Segundo esses dados é possível observar que muitas vezes a distância necessária que o paciente precisa percorrer até chegar em centros urbanos, principal local de atuação dos médicos, pode ocasionar dificuldade do acesso a diagnósticos e tratamentos.

Levando em consideração as dificuldades mencionadas e aproveitando-se da evolução tecnológica dos últimos anos, percebe-se a importância do desenvolvimento de métodos computacionais que possam permitir a detecção de patologias da visão. Tais métodos podem ser utilizados para fornecer acesso básico a saúde da visão para uma maior parcela da população, auxiliar em processos de triagem e servir de suporte para médicos.

Este trabalho descreve um método para diagnosticar o estrabismo. O método proposto baseia-se no emprego da rede neural convolucional YOLOv5 para detectar os olhos e o oclutor durante o processamento de vídeos do *Cover test* alternado de voluntários adquiridos em consultório médico e em escolas públicas da cidade de São Luís no estado do Maranhão, Brasil. Finalmente, os resultados do método proposto são comparados com as medições de um médico especializado em estrabismo. Para esse propósito foram realizadas tarefas como: definição de uma metodologia de aquisição de vídeo, a aquisição de um conjunto de dados, e emprego de um método de aprendizado profundo para auxiliar na detecção e diagnóstico de estrabismo. Essas tarefas possibilitaram a construção de um método com boa precisão em ambientes distintos, como consultório médico e ambiente escolar. Essa flexibilidade pode permitir sua utilização em triagens, aliviando o sistema

¹ O teste de Hirschberg consiste na análise da posição relativa do reflexo corneano, através da iluminação simultânea binocular. Permite identificar a presença de desvios oculares manifestos (heterotropia)

² O Teste de Lancaster é um procedimento utilizado para identificar quais músculos da estrutura ocular estão enfraquecidos ou paralisados.

saúde pública.

Considerando o que já foi mostrado, o método proposto neste trabalho se difere de outros trabalhos presentes na literatura em alguns pontos. No primeiro ponto, destaca-se a utilização de uma rede neural para detecção de múltiplos objetos na etapa de detecção de olhos e oclutor, permitindo uma maior robustez nessa etapa e possibilitando uma maior flexibilidade na aquisição dos vídeos do exame. Em segundo lugar, o método é avaliado em duas bases de exames do *Cover test* alternado com bastante diferença entre si, uma foi adquirida em consultório médico e trata-se de um ambiente mais controlado, enquanto a outra foi obtida em escolas públicas com protocolo de aquisição menos rígido. Por último, outro ponto de distinção é a comparação dos resultados com outro trabalho da literatura e com diagnóstico do médico especialista.

1.1 Objetivo Geral

O objetivo do trabalho é utilizar uma rede neural convolucional para detecção de múltiplos objetos, no desenvolvimento de uma metodologia para detectar e diagnosticar estrabismo em vídeos do exame *Cover test* alternado.

1.2 Objetivos específicos

Para alcançar o objetivo geral pretendido é necessário desenvolver os seguintes objetivos específicos:

- Construir uma base de dados com vídeos do exame *Cover test*;
- Definir uma arquitetura de rede neural convolucional (CNN) para rastrear múltiplos objetos;
- Empregar modelo de CNN para a detecção dos olhos e oclutor;
- Desenvolver um método para medir o desvio ocular em vídeos do *Cover test*;
- Comparar resultados do método com os dados do especialista e outros trabalhos da literatura.

1.3 Organização da Dissertação

O trabalho está organizado como segue: O [Capítulo 2](#) apresenta alguns dos principais trabalhos relacionados ao tema abordado. Nele, cada trabalho é descrito com foco em suas

principais características e contribuições. Em seguida o [Capítulo 3](#) aborda os principais conceitos sobre o olho humano, o estrabismo e redes neurais, relevantes para a compreensão deste trabalho. No [Capítulo 4](#) aborda-se a metodologia proposta para o diagnóstico do estrabismo em vídeos do exame *cover test*. Também é apresentado os materiais utilizados no processo de construção do método. Já no [Capítulo 5](#) são apresentados os resultados alcançados, seguido de uma discussão sobre os mesmos. Por último, o [Capítulo 6](#) descreve as conclusões obtidas deste trabalho, destacando as contribuições geradas por esta pesquisa e os trabalhos futuros que podem ser explorados.

2 TRABALHOS RELACIONADOS

O problema de detecção e diagnóstico automático de estrabismo não é um problema novo e por isso na literatura é possível encontrar trabalhos relacionados ao tema. Com o objetivo de filtrar os trabalhos mais relevantes sobre o assunto e conhecer as abordagens adotadas anteriormente junto de seus pontos de melhoria, foi realizada uma Revisão Sistemática da Literatura (RSL).

2.1 Revisão Sistemática da Literatura

A elaboração da revisão sistemática da literatura (RSL) sobre métodos computacionais aplicados ao estrabismo deu-se por meio do uso da plataforma *Google Scholar*, da qual foram filtrados artigos científicos relacionados aos termos chaves: “Strabismus diagnosis”, “Diagnóstico automático de estrabismo”, “Strabismus Detection”, “Detecção automática de estrabismo”, “Strabismus Recognition” e “Strabismus Measurement”. Foram considerados apenas artigos publicados em um intervalo específico de tempo que compreende os anos de 2010 ao ano de 2021 e que foram escritos nos idiomas utilizados para filtragem, nesse caso Inglês e Português.

Para a realização de uma revisão sistemática da literatura, é preciso a elaboração de questões de pesquisa(QP), as quais serviram para sistematizar a validação de publicações que satisfaçam o objetivo do estudo. Foram elas:

- QP1. O estudo realiza detecção ou diagnóstico?
- QP2. O estudo utiliza alguma informação como vídeos ou imagens digitais?
- QP3. O estudo construiu ou utiliza alguma ferramenta que automatize o processo de detecção ou diagnóstico?

Com o objetivo de verificar a relevância dos artigos, que foram considerados após a aplicação das questões de pesquisa, os trabalhos passaram por uma avaliação conforme o modelo empregado em [Maciel, Matias e Sarinho \(2019\)](#). O objetivo das perguntas de qualidade é avaliar características dos métodos que se considera mais relevantes como: a automatização do diagnóstico, emprego de técnicas de processamento de imagem ou vídeo, emprego de técnicas de aprendizado de máquina, o cálculo do desvio e por fim o custo da implementação e uso do método desenvolvido no trabalho. Para isso, as seguintes perguntas foram feitas considerando o conteúdo:

- PQ1. O estudo utilizou, construiu ou propôs alguma ferramenta ou método para automatizar a detecção, o diagnóstico?
- PQ2. O estudo utiliza alguma técnica de processamento de imagem ou vídeo para a detecção, ou diagnóstico?
- PQ3. O estudo utiliza alguma técnica de aprendizado de máquina profundo para a detecção, ou diagnóstico?
- PQ4. Além de diagnosticar o estrabismo, o estudo calcula o desvio?
- PQ5. O método ou ferramenta proposta pelo estudo tem custo de implementação acessível?

A primeira pergunta (PQ1) verifica se o trabalho desenvolvido, de fato, automatiza o processo de detecção, diagnóstico ou planejamento cirúrgico do estrabismo, visto que essa característica pode auxiliar os profissionais especializados possibilitando-os atender uma maior demanda de pacientes. A pergunta 2 (PQ2) averigua se o trabalho emprega técnicas de processamento de imagem ou vídeo na resolução do problema, considerando que afeta características físicas da face. A terceira pergunta (PQ3) complementa a segunda, já que na última década grandes avanços na área de aprendizado de máquina foram alcançados, permitindo que novas características e formas de interpretação de dados possam ser feitas. Com a quarta pergunta (PQ4) busca-se trabalhos que permitam, além de diagnosticar o paciente como saudável ou estrábico, medir o desvio para avaliar a gravidade e planejar futuras ações. A última pergunta (PQ5) contempla trabalhos em que o custo de implementação ou uso do método proposto seja acessível, permitindo que mais pessoas possam ter acesso aos benefícios da pesquisa desenvolvida.

Cada pergunta foi respondida com Sim (S) equivalente a 1 ponto, Parcial (P) igual a 0.5 pontos ou Não (N) igual a 0 pontos. Dessa forma, os melhores artigos foram aqueles em que a nota mais se aproximou de 5.

2.1.1 Resultados da RSL

Para o desenvolvimento da revisão foi realizada uma busca considerando os critérios de inclusão e exclusão mostrados anteriormente nas perguntas de pesquisa, compreendendo desta forma a 1ª fase. A 2ª fase, é o processo de busca pelos termos chaves, aos quais foram mostrados no começo da seção, no título e resumo dos trabalhos. Por último, na 3ª fase, são considerados apenas artigos que contribuem diretamente com a detecção, diagnóstico automático de estrabismo, aplicando as perguntas QP1 à QP3 discutidas.

Desta maneira, foram selecionados 3.601 artigos após a 1ª fase, 48 artigos após aplicação do filtro da 2ª fase e 16 após filtragem da última fase. Como resultado, os artigos

remanescentes e fortemente relacionados a detecção, planejamento cirúrgico ou diagnóstico automático de estrabismo são apresentados na Tabela 1, em que são mostrados o título, ano de publicação e a quantidade de citações até o momento do desenvolvimento desta revisão.

Tabela 1 – Trabalhos selecionados após 3ª fase.

ID	Ano	Título	Citações
T1	2012	Computational methodology for automatic detection of strabismus in digital images through Hirschberg test (ALMEIDA et al., 2012)	40
T2	2015	Computer-Aided Methodology for Syndromic Strabismus Diagnosis (ALMEIDA et al., 2015)	24
T3	2015	The automated diagnosis of strabismus using an infrared camera (SEO et al., 2015)	9
T4	2017	Automatic diagnosis of strabismus in digital videos through cover test (VALENTE et al., 2017)	21
T5	2017	Strabismus Measurements with Novel Video Goggles (WEBER et al., 2017)	14
T6	2018	Automated Strabismus Detection for Telemedicine Applications (LU et al., 2018)	12
T7	2018	An Automatic Stimulus and Synchronous Tracking System for Strabismus Assessment based on Cover Test (ZHENG et al., 2018)	6
T8	2019	Automatic ocular alignment evaluation for strabismus detection using U-NET and ResNet Networks (SIMOES et al., 2019b)	-
T9	2019	Avaliação automática do alinhamento ocular da detecção de estrabismo usando índice de diversidade filogenética e máquina de vetores de suporte (SIMOES et al., 2019a)	-
T10	2019	Quantitative measurement of horizontal strabismus with digital photography (DERICIOĞLU; ÇERMAN, 2019)	3
T11	2020	Virtual reality-based measurement of ocular deviation in strabismus (MIAO et al., 2020)	14
T12	2020	An Image-Based Method for Measuring Strabismus in Virtual Reality (MEHRINGER et al., 2020)	-
T13	2021	Detection of Referable Horizontal Strabismus in Children's Primary Gaze Photographs Using Deep Learning (ZHENG et al., 2021)	5
T14	2021	Range Detection of Strabismus based on the Distance and Coordinates of the Iris (ZOLKIFLI et al., 2021)	1

Como mencionado anteriormente, os artigos selecionados após a 3ª fase foram avaliados pela qualidade e relevância do conteúdo considerando o objetivo da revisão. Para isso, os artigos foram lidos e examinados conforme as perguntas de qualidade PQ1 à PQ5. A avaliação de cada trabalho é apresentada na Tabela 2, em que é possível observar a pontuação alcançada e a resposta para cada pergunta.

De forma a organizar a apresentação dos artigos selecionados foram considerados os principais objetivos propostos e resultados alcançados. Primeiramente são discutidos os artigos cujo o objetivo principal é detectar o estrabismo, através de classificação do paciente como saudável ou estrábico (Seção 2.2). Logo em seguida, são discutidos os trabalhos que realizam o diagnóstico e medição do desvio do alinhamento ocular (Seção 2.3).

Tabela 2 – Avaliação de qualidade do trabalhos

Trabalho	PQ1	PQ2	PQ3	PQ4	PQ5	Pontos
T1	S	S	N	S	S	4
T2	S	S	N	S	S	4
T3	S	S	N	P	S	3.5
T4	S	S	N	S	S	4
T5	S	N	N	S	N	2
T6	S	S	S	N	S	4
T7	S	S	N	N	P	2
T8	S	S	S	N	S	4
T9	S	S	S	N	S	4
T10	S	S	N	S	P	3.5
T11	S	S	N	S	N	3
T12	S	S	N	S	N	3
T13	S	S	S	N	S	4
T14	S	S	N	P	S	3.5

2.2 Detecção de estrabismo

Alguns dos trabalhos selecionados têm como objetivo identificar se o paciente é estrábico ou não. Esses trabalhos são importantes mesmo não apresentando um diagnóstico que informe a medida do desvio ou planeje uma intervenção cirúrgica, por apresentarem métodos capazes de realizar triagens em grande quantidades de pacientes.

O trabalho de [Almeida et al. \(2012\)](#) aplica processamento de imagem para realizar a detecção de estrabismo em imagens do teste de Hirschberg. A metodologia é dividida em quatro partes, na qual, primeiramente, a região dos olhos é selecionada através do corte da imagem para descartar as regiões irrelevantes. A segunda etapa utiliza equalização do histograma, extração de características e *Support Vector Machine* (SVM) para classificar as regiões que melhor representam os olhos. O limbo e o brilho são encontrados através da detecção de bordas e de círculos, respectivamente. Por fim, é medida a distância entre o brilho e o centro do limbo para diagnosticar se o paciente é estrábico. O trabalho obteve uma acurácia de 93,8% para identificar pacientes doentes nos casos que a detecção dos olhos obteve sucesso.

[Seo et al. \(2015\)](#) apresentaram uma automação do exame para a detecção de estrabismo, através do acionamento de obturadores presentes em um óculos e a aquisição de diversas imagens com o auxílio de uma câmera infra-vermelha. As imagens resultantes são processadas e o estrabismo é detectado através do cálculo da diferença entre o centro do olho e a posição do brilho. Como resultado, o trabalho afirma que o erro médio, entre a medição e o valor real, é de 1,66 mm. Contudo, o método seria melhor explorado caso a medição do desvio fosse empregada para a geração de um diagnóstico mais completo, além da medida ser informada em Δ (Dioptrias prismáticas), unidade de medida comumente utilizada pelos especialistas. Outro fator importante é que as ferramentas utilizadas podem não ser de fácil acessibilidade.

O trabalho apresentado por (LU et al., 2018) faz a utilização de redes neurais convolucionais para a segmentação da região dos olhos e classificação do paciente como estrábico ou não. A metodologia é dividida em duas etapas distintas, a primeira utiliza *Region-based Fully Convolutional Networks* (R-FCN) para segmentar a região dos olhos e a segunda utiliza uma CNN para classificá-lo como estrábico ou normal. Dentre os resultados foram alcançados 95% de sensibilidade, 94% de especificidade e 95% de acurácia. Como principais contribuições pode-se ressaltar o desenvolvimento de um método viável e pouco parametrizado, no qual facilita a reprodução. Porém, os requisitos para processamento das imagens como, por exemplo, a placa gráfica podem ser considerados pouco acessíveis.

Recentemente, o trabalho de Zheng et al. (2018) dá continuidade ao método mostrado em Lu et al. (2018). Com o objetivo de melhorar os resultados obtidos pela metodologia anterior, os autores utilizaram uma base de dados com uma quantidade maior de imagens. Junto ao aumento da base de dados empregaram no treinamento da R-FCN, utilizada na primeira etapa do método para identificar a região dos olhos. Outra modificação importante na metodologia, foi a substituição da CNN utilizada para classificar a região segmentada. No primeiro momento foi utilizada uma arquitetura mais simples, porém foi substituída por uma versão pré-treinada da Inception-V3, uma rede neural mais robusta, que permite uma melhor classificação.

Como resultado das modificações apresentadas por Zheng et al. (2018), vale ressaltar a melhora nos resultados, uma vez que a AUC alcançou 0,993, bem como, aplicou um processo de validação mais robusto, no treinamento, e comparou os resultados do método com o diagnóstico de três oftalmologistas, os quais apresentam uma taxa de erro maior quando comparado ao método proposto. Outra ação importante foi a substituição da rede utilizada na fase de classificação por um modelo pré-treinado, que permitiu a redução do custo do processo, tendo em vista a simplificação do hardware necessário.

A pesquisa exposta por (ZOLKIFLI et al., 2021) propõe um sistema de detecção semiautomática baseado na distância e nas coordenadas da íris, com o objetivo de ajudar a reduzir o tempo de triagem para o oftalmologista diagnosticar o estrabismo. É dividido em três etapas: pré-processamento para melhorar a imagem original, localização da íris e classificação do paciente. O pré-processamento consiste na extração do canal vermelho da imagem e suavização. A localização da íris é feita com limiares e cortes de regiões de interesse. A classificação é feita calculando a distância do centro da íris, considerando intervalos de valores de pacientes normais. O método proposto tem como benefício a simplicidade e rapidez, contudo, é sensível a fatores externos como iluminação e protocolo de aquisição, devido a etapas como a de aplicação de limiares. Outro fator que faz falta no trabalho é a validação dos resultados junto com um especialista em estrabismo.

Conforme foi mencionado no início dessa seção, os trabalhos discutidos têm como objetivo principal classificar os pacientes em saudáveis ou estrábitos. Vale ressaltar que,

mesmo os trabalhos não apresentando uma medição do desvio ou apenas lidarem com estrabismos horizontais em sua maioria, os resultados alcançam taxas elevadas de acerto, mesmo quando comparados com os especialistas. Dessa forma, os métodos podem auxiliar os oftalmologistas nas triagens, auxiliando-os na detecção de desvios oculares.

2.3 Diagnóstico de Estrabismo

Esta seção aborda os artigos selecionados que têm, dentre seus objetivos ou etapas da metodologia, a medição do desvio e classificação do tipo do desvio, diferente dos trabalhos da Seção 2.2, apresentam diagnósticos mais completos, permitindo que o especialista tenha mais informação a respeito do grau de estrabismo de um paciente.

Almeida et al. (2015) melhoraram seu trabalho anterior, discutido no início da Seção 2.2, no qual realizava apenas a detecção do estrabismo. Dessa forma, foram propostas algumas modificações na etapa de localização dos olhos com a utilização de funções geoestatísticas além de realizar a medição do desvio mais precisa, através de experimentos sobre o tamanho aproximado do limbo, alcançando uma acurácia de 98,5%. O trabalho contribui com um método capaz de realizar um diagnóstico completo, desde a identificação de pacientes doentes até uma aproximação do valor do desvio. A maior dificuldade que pode ser observada no método é sua grande quantidade de parâmetros o que pode dificultar a sua utilização em outras bases de imagens.

Valente et al. (2017) realizam o diagnóstico do estrabismo através do processamento de vídeos digitais do *Cover-test*. O pré-processamento utiliza morfologia matemática para a localização da pupila. A equalização do histograma, o detector de bordas e a transformação de Hough localizam o limbo. Por fim, a subtração de fundo é feita para detectar o ocluser, pois o *Cover-test* baseia-se em seu movimento. O trabalho contribui com uma metodologia para resolver problemas que dependem do rastreamento ocular, um método de baixo custo quando se considera os hardwares necessários. Além disso, o método pode ser aplicado a qualquer pessoa. Como resultados, obteve uma precisão de 87%, especificidade de 100%, sensibilidade de 80% e realiza uma medição de desvio horizontais considerando a mesma unidade de medida dos especialistas, bem como permite o diagnóstico do estrabismo latente. Os pontos negativos observados são a grande quantidade de parâmetros, o que pode dificultar a adaptação para outras bases de dados, e a grande dependência de fatores externos, tais como, protocolo rígido de aquisição dos vídeos, a dependência da iluminação do ambiente e a pequena quantidade de vídeos utilizada.

Uma abordagem que faz uso de óculos para a detecção do estrabismo é explorada por Weber et al. (2017). Os participantes foram testados com óculos de vídeo infravermelho com projeção de alvo a laser embutido. Houve boa concordância entre os resultados obtidos

com o óculos e o padrão de referência existente, o teste de tela de Hess Lancaster. Nesse teste são mostrados pontos em uma tela e o paciente deve segui-los com o olhar. As medidas de desvio horizontal e vertical apresentaram uma correlação intraclasse horizontal de 95%. A metodologia apresentada se mostrou simples, rápida e precisa na medição de desvios oculares. Contudo, a aquisição do óculos e da tela necessária para a execução do exame pode torná-lo pouco acessível. Outro fator que poderia ter sido mais explorado é a utilização do cálculo do desvio não apenas para comparação com o valor de referência, mas, na geração do laudo.

Zheng et al. (2018) criaram um módulo de estímulo automático, um conjunto de hardware e software para a execução e avaliação do teste de cobertura. O hardware é composto por um monitor, em que pontos são observados, um *slider*, o qual move o ocluser, um controlador para o *slider* e uma estação de trabalho para o processamento do vídeo. O software realiza a localização e acompanhamento do centro da pupila para a análise do movimento dos olhos. O resultado do desvio medido é entre 0,5 unidades e 1 unidade. A principal contribuição do método é a automatização completa do teste de cobertura, que pode ser executado por pessoas após um breve treinamento, não sendo necessário um especialista da área. Contudo, uma validação dos resultados do método junto com um especialista ou comparando com diagnósticos anteriores não é apresentado. Outro fator que pode comprometer o funcionamento do método é a inacessibilidade dos equipamentos devido ao seu valor e a mão de obra especializada para a manutenção dos mesmos.

Nos trabalhos (SIMOES et al., 2019a) e (SIMOES et al., 2019b), de mesma autoria, são propostos métodos bem similares para o diagnóstico do estrabismo em imagens do teste de Hirschberg. A metodologia é capaz de classificar o tipo de estrabismo por meio da avaliação do alinhamento ocular em relação aos cantos dos olhos. As principais etapas da metodologia explorada nos trabalhos são: aquisição de imagens, localização dos olhos, pré-processamento, segmentação da esclera, localização do limbo e detecção do estrabismo. Em um primeiro momento a segmentação da esclera foi realizada dividindo a região dos olhos em sub-regiões e extraíndo características para classificar as regiões em esclera e não esclera utilizando o *Support Vector Machine* (SVM). A abordagem apresentada pelos autores se provou insuficiente e logo em seguida foi substituída pelo uso da U-NET. A modificação realizada no método permitiu que o trabalho alcançasse 95,8% de sensibilidade, 100% de especificidade e uma acurácia de 96,6%. Contudo, o trabalho poderia ter utilizado o cálculo da distância do centro da pupila até o brilho, aplicado na etapa de detecção do estrabismo, para estimar o desvio do paciente e, dessa forma, proporcionar um diagnóstico mais completo.

Em (DERICIOĞLU; ÇERMAN, 2019) foi desenvolvido um método para calcular o ângulo do olhar por meio de fotografias de pacientes com estrabismo para determinar sua validade e confiabilidade. O método faz uso de um aplicativo para celular capaz de

medir as seguintes distâncias em pixels: do centro geométrico para a córnea ao reflexo da luz, diâmetro da córnea e distância interpupilar. Com esses valores é calculado o ângulo do desvio permitindo, dessa maneira, o diagnóstico do paciente. Como resultados, é apresentado que o método alcançou um erro máximo de $6,1\Delta$ mostrando que o ângulo medido se aproxima do real. Contudo, o método é limitado ao diagnóstico apenas do estrabismo manifesto e poderia ter sido empregado técnicas para a detecção dos olhos e limbo já disponíveis na literatura, assim ficaria menos vulnerável a erros, devido às marcações manuais.

Outros trabalhos como (MIAO et al., 2020) e (MEHRINGER et al., 2020) fazem uso de um *Head-mounted display*, óculos de realidade aumentada, para detectar e medir o desvio do estrabismo. Por se tratarem de equipamentos que ficam bem próximos, permitem controlar as imagens expostas aos olhos de forma individual, bem como, simular o *Cover test* alternado (MIAO et al., 2020) e o *Prism Cover test* (MEHRINGER et al., 2020). Os resultados dos trabalhos apresentam erros máximos que variam de 5 a 10Δ , mostrando a precisão alcançada. Embora haja diversas vantagens, dois problemas principais encontradas nos métodos são: o valor do óculos de realidade virtual que, infelizmente, não é acessível em sua grande maioria e o hardware necessário para o processamento dos cenários virtuais.

2.4 Considerações

Diferentemente dos trabalho citados na Seção 2.2 que buscam apenas informar se o paciente é estrábico ou não, este trabalho tem como finalidade além de detectar, realizar o diagnóstico do paciente através da cálculo dos desvios horizontais e verticais. Contudo, o método proposto também se diferencia dos trabalhos apresentados na Seção 2.3, por tratar-se de uma abordagem que busca integrar várias características para realizar o diagnóstico, tais como: utilização de uma rede neural convolucional para a detecção dos olhos e oclusor em vídeos do exame *Cover Test*, utilização de hardwares mais acessíveis - diferente de alguns trabalhos que necessitam de hardwares específicos, um protocolo de aquisição menos rígido, um método que seja mais robusto e a aplicação em uma base de vídeos mais ampla.

Alguns dos trabalhos mencionados nessa seção, principalmente os que realizam a medição do desvio, utilizam o cálculo da diferença entre as posições do centro do limbo antes e depois de ser ocluído. Contudo, um dos principais problemas encontrados nessa abordagem é a dependência de uma boa localização das regiões englobadas pelo limbo. A exemplo do trabalho apresentado por Valente et al. (2017), o qual dedica uma etapa de seu método apenas para a localização do limbo, uma etapa extensa e muito parametrizada, que acaba dependente de vários fatores externos e um protocolo mais rígido para aquisição dos vídeos. Outros trabalhos ((WEBER et al., 2017),(ZHENG et al., 2018),(DERICIOĞLU;

(ÇERMAN, 2019), (MIAO et al., 2020), (MEHRINGER et al., 2020)) também mostrados na Seção 2.3, buscam minimizar a dificuldade desta etapa através de marcações manuais, construção de hardwares específicos ou até mesmo óculos de realidade virtual, o que permite um melhor controle do cenário e conseqüentemente uma determinação dessa região de forma precisa, porem acabam encarecendo o método e dificultando a acessibilidade. Um resumo dos pontos e considerações levantados é mostrado na 3.

Tabela 3 – Resumo dos trabalho relacionados

Autor	Exame base	Controle do Ambiente	Custo	Objetivo	Consideração
Almeida et al. (2012)	Hirschberg	Médio	Baixo	Detectar	Estrabismo Manifesto
Almeida et al. (2015)	Hirschberg	Médio	Baixo	Diagnósticar/ Medição do desvio	Estrabismo manifesto Muito parametrizado
Seo et al. (2015)	Cover test	Médio	Alto	Detectar	Óculos infra-vermelho
Valente et al. (2017)	Cover test Alternado	Alto	Baixo	Diagnósticar/ Medição do desvio	Muito parametrizado
Weber et al. (2017)	Hess Lancaster	Alto	Alto	Diagnósticar	Óculos infra-vermelho/Tela especial Marcação laser
Lu et al. (2018)	-	-	Baixo	Detectar	Estrabismo manifesto
Zheng et al. (2018)	-	-	Baixo	Detectar	Estrabismo manifesto
Simoès et al. (2019a)	Hirschberg	Médio	Baixo	Diagnósticar	Estrabismo manifesto
Simoès et al. (2019b)	Hirschberg	Médio	Baixo	Diagnósticar	Estrabismo manifesto
Dericioğlu e Çerman (2019)	-	Médio	Baixo	Diagnósticar/ Medição do desvio	Estrabismo Manifesto Marcação manual
Miao et al. (2020)	Cover test	Médio	Alto	Diagnósticar/ Medição do desvio	Óculos RV
Mehringner et al. (2020)	Prism Cover test	Médio	Alto	Diagnósticar/ Medição do desvio	Óculos RV
ZOLKIFLI et al. (2021)	-	Baixo	Baixo	Deteção	Não valida com especialista
Neste trabalho	Cover test alternado	Baixo	Baixo	Diagnósticar/ Medição do desvio	Validação com especialista

3 FUNDAMENTAÇÃO TEÓRICA

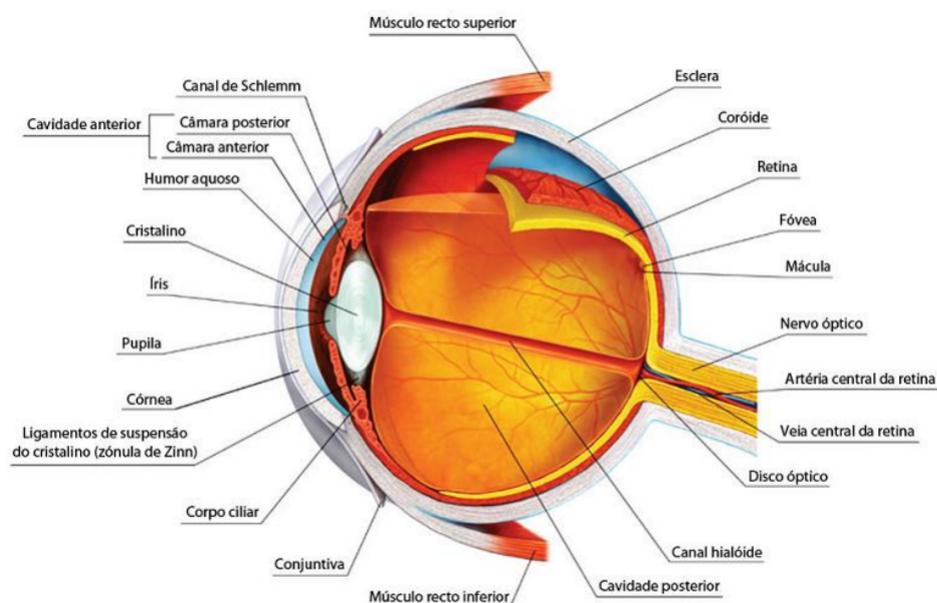
O seguinte capítulo descreve técnicas e conceitos importantes para a compreensão do método proposto. Primeiro, é necessário discorrer sobre o olho humano, órgão que pode ser afetado por patologias oftálmicas. Logo após, é explanado sobre o estrabismo, doença que o método busca diagnosticar. Em seguida, o *cover test* será explicado, um dos exames utilizados para a detecção do estrabismo e o exame executado durante os vídeos adquiridos para o desenvolvimento deste trabalho. Nos tópicos seguintes, são abordados definições de redes neurais, partindo do conhecimento mais amplo, passando por redes neurais convolucionais e detecção de objetos. Por fim, é descrita a família de redes neurais conhecida como YOLO e sua última versão utilizada nesse trabalho, a YOLOv5.

3.1 O Olho humano

O olho humano é o órgão responsável pela fotorrecepção, processo pelo qual a energia luminosa produz alterações em estruturas especializadas da retina e as transmite para o cérebro (BAIÃO, 2013). O olho possui uma forma esférica com aproximadamente 24 mm de diâmetro, é uma estrutura complexa e fica suspenso na cavidade óssea por músculos que controlam seus movimentos. A estrutura detalhada do olho é apresentada na Figura 1. Dentre os componentes apresentados, é possível observar os que compõem as três camadas oculares, também chamadas de túnicas, sendo elas: a camada externa (túnica fibrosa), camada média (túnica vascular) e a camada interna (túnica interna) (HEALTHENGINE, 2007).

A camada externa do olho é composta pela esclera, parte branca, a qual cobre maior parte da superfície do olho, essa parte é rica em vasos sanguíneos e fornece a fixação para os músculos externos. Além da esclera, a córnea também compõem a primeira camada do olho e é o primeiro componente refrativo do olho. A segunda camada ocular é a média, dentre sua composição se destacam o corpo ciliar e a íris. O corpo ciliar tem formato de anel, controla a pressão ocular, mantém o formato esférico do globo e produz fluidos intraoculares como o humor aquoso. Por sua vez, a íris é um músculo fino e ajustável que controla a quantidade de luz que alcança a retina. O controle de luz, é realizado através da modificação do diâmetro da pupila (HEALTHENGINE, 2007). A última camada do olho, túnica interna, contém os componentes neurais. Esses componentes, são células fotorreceptoras chamadas de bastonetes e cones da retina, responsáveis por codificar a informação visual em impulsos elétricos que são transmitidos ao longo do nervo óptico para o cérebro.

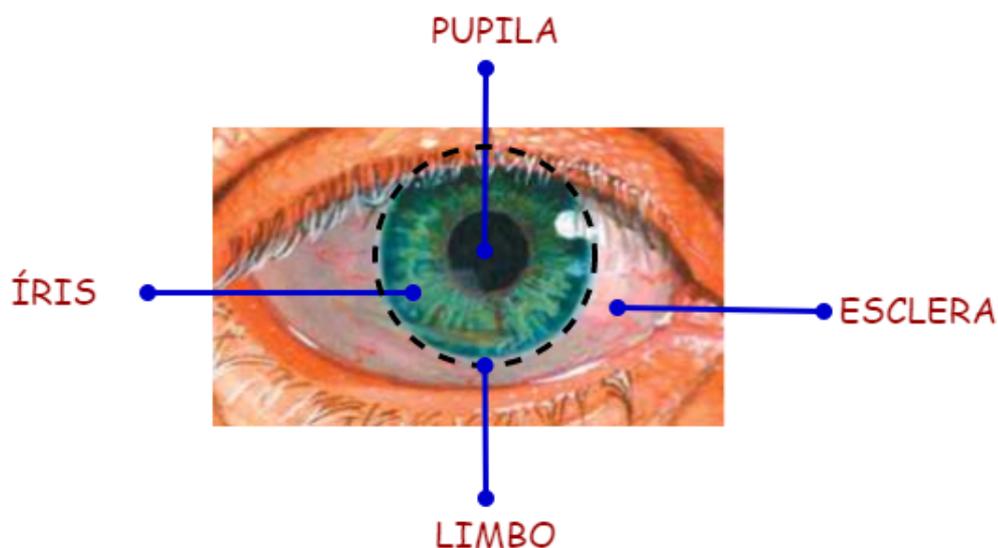
Figura 1 – Estrutura detalhada do olho humano.



Fonte: Baião (2013).

Para a execução do método proposto, foram consideradas as estruturas visíveis do olho humano apontadas na Figura 2. Vale ressaltar, que o centro da pupila e a região fronteira entre a esclera e a íris, denominada de limbo, são amplamente exploradas pelo método que será detalhado no decorrer deste trabalho (Seções 4.2 e 4.3).

Figura 2 – Estruturas visíveis do olho.



Fonte: turner (2012).

3.1.1 Visão Binocular

Como apresentado na seção anterior, o olho humano é complexo, capaz de captar imagens, interpretá-las e codificá-las em impulsos elétricos para que possam ser enviadas ao cérebro. Contudo, o ser humano possui dois olhos, cada olho capta uma imagem e possui um campo de visão. O campo de visão de um olho pode ser abstraído como um cone, no qual o seu vértice está localizado no olho e sua base é projetada para o ambiente. O espaço comum dos campos de visão dos dois olhos é chamado de campo binocular da visão.

Após cada olho adquirir a imagem os sinais nervosos se deslocam de cada um, junto com o nervo óptico correspondente, para a parte posterior do cérebro, local em que os dois nervos ópticos se encontram, conforme mostrado na Figura 3. Lá, o nervo óptico de cada olho se divide, o lado direito do cérebro recebe informação dos dois nervos ópticos para o campo esquerdo da visão e o lado esquerdo do cérebro recebe informação dos dois nervos ópticos para o campo direito da visão, esse processo é conhecido como visão binocular.

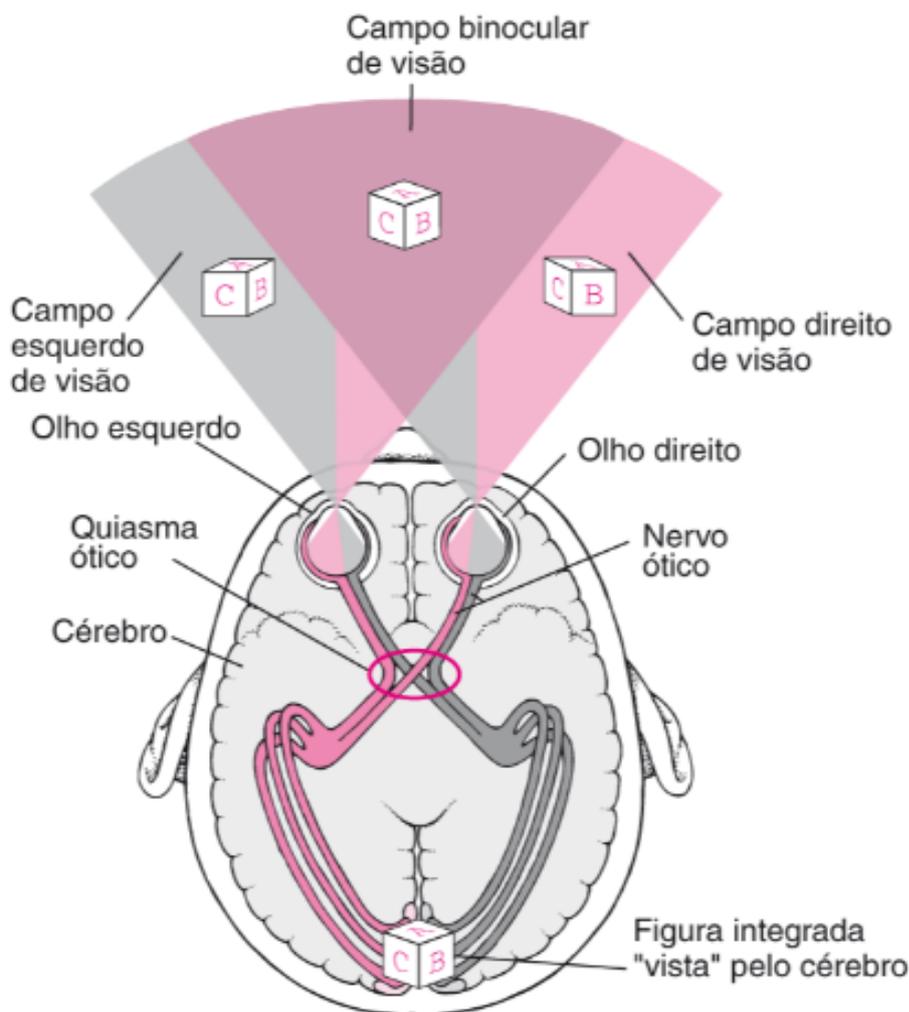
A visão binocular é um processo complexo e de extrema importância. Através dele o ser humano tem a capacidade de perceber a profundidade, característica indispensável para uma boa qualidade de vida, e para execução de tarefas mais seguras e precisas.

3.2 O Estrabismo

O estrabismo, é de uma condição em que a linha de visão de um dos olhos não consegue fixar no objeto alvo que os dois olhos estão visualizando, ou seja, enquanto um dos olhos fixa o olhar no objeto, a linha de visão do outro está para uma direção diferente (LEITMAN, 2021). Além disso, o estrabismo pode ser classificado considerando a forma de manifestação. Quando os eixos da visão não estão alinhados formando a visão binocular, ou seja, o desalinhamento pode ser percebido apenas olhando para o paciente, o estrabismo é classificado de manifesto ou tropia. Por outro lado, quando os eixos de visão conseguem formar a visão binocular e o desalinhamento só é percebido em situações específicas, essa forma de estrabismo é conhecida como latente ou foria (COSTA, 2017). Vale ressaltar que o estrabismo latente é a forma mais difícil de ser diagnosticada, pois necessita de exames específicos.

O estrabismo também pode ser classificado de acordo com as possíveis direções do desalinhamento ocular. Conforme mostrado na Figura 4, são quatro as possíveis direções de desvio, sendo elas: (A) desvio horizontal para fora (Exotropia), (B) desvio horizontal para dentro (Esotropia), (C) desvio vertical para cima (Hipertropia) e (D) desvio vertical para baixo (Hipotropia).

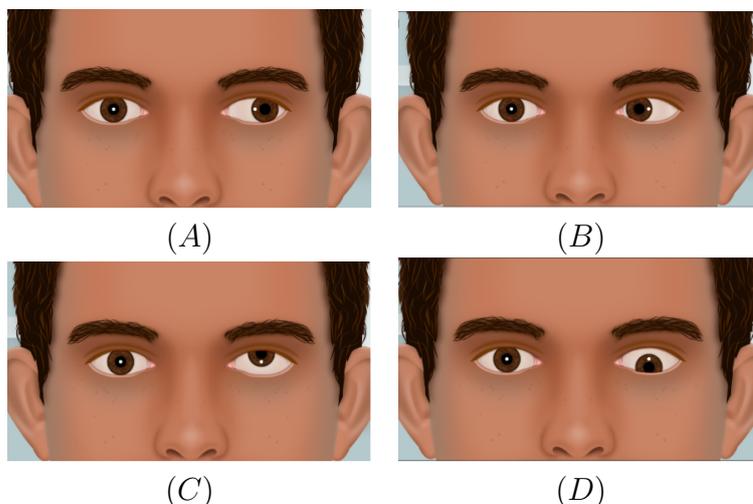
Figura 3 – Exemplificação da visão binocular.



Fonte: Garrity (2019).

Afetando a visão, um dos principais sentidos do ser humano, o estrabismo afeta diretamente a qualidade de vida da pessoa acometida pelo distúrbio. O distúrbio não prejudica apenas a parte física, mas aspectos psicossociais. Quando ela se manifesta ainda na infância, as crianças podem ter maior dificuldade de aprendizado, ser alvos de bullying, dificuldade em socializar, fazer amizades e até desenvolver outras doenças oculares (BOMMIREDDY; TAYLOR; CLARKE, 2020). Caso o estrabismo não seja tratado ainda na infância quando se tem o diagnóstico precoce ou em casos que o distúrbio se manifesta na vida adulta, o estrabismo pode ocasionar problemas graves na visão, prejudicar aspectos como a personalidade, baixo auto-estima, impedir a execução de tarefas, principalmente as que dependem da noção de profundidade e por fim, dificultar a empregabilidade da pessoa afetada (HATT et al., 2007).

Figura 4 – Estrabismo conforme direção do desvio. (A) Exotropia (B) Esotropia (C) Hipertropia (D) Hipotropia.



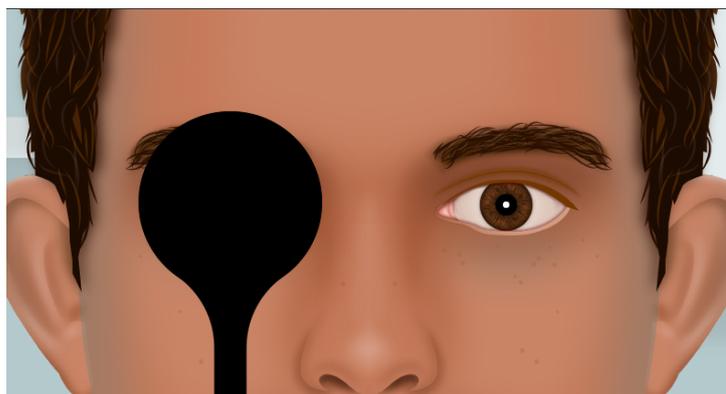
Fonte: Adaptado de Faruk (2015).

3.3 Cover Test

O *Cover test* é um exame ocular comumente utilizado para a detecção e diagnóstico do estrabismo. Nele é exigido que o paciente fixe os olhos em um alvo e logo após um dos olhos é coberto com um objeto chamado de ocluser. Uma dificuldade para a execução do exame, é a necessidade de fixação do olhar em um ponto fixo por um determinado tempo, o que pode ser difícil para crianças e pessoas com hiperatividade. Movimentos bruscos, comuns a esse público, podem influenciar negativamente no resultado. Para contornar esses problemas, podem ser adotadas táticas como a inserção de objetos atrativos que capturem a atenção de crianças e executar o exame com pausas para deixá-las mais confortáveis (TRABOULSI; UTZ, 2016).

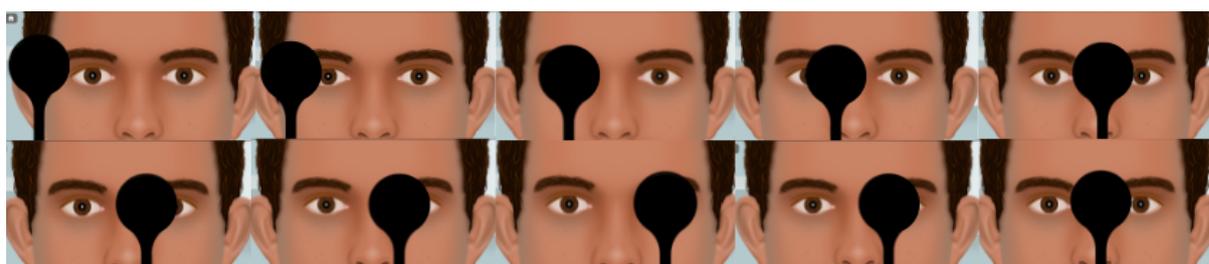
O estrabismo, como explicado na seção anterior, pode causar heterotropias ou heteroforias. Dessa forma, o *cover test* é capaz de detectar e diagnosticar o estrabismo em suas duas formas, porém, é necessário algumas modificações e adaptações em sua execução. Na primeira situação de desvios manifestos, utiliza-se o *cover test* em sua versão mais simples, conforme mostrado na Figura 5. Para a realização do exame, um olho é coberto e observa-se o comportamento do olho oposto para averiguar se existe algum tipo de movimento involuntário, em seguida, o ocluser é removido. Espera-se o paciente recuperar a visão binocular e então repete-se o exame, porém, agora ocluindo o outro olho. Caso exista movimento involuntário no olho que está sendo coberto ou descoberto, possivelmente é indicação da presença de desvio no olho oposto. Esse processo de mudança no estado dos olhos, é a tentativa do cérebro compensar o desalinhamento e a busca para fixação no alvo.

No segundo caso, para detecção de heteroforias, é necessário uma execução diferente

Figura 5 – Execução do *Cover test*.

Fonte: Adaptado de Faruk (2015)

do *cover test*. Essa forma de execução do exame é mostrada na Figura 6 e dá origem ao exame chamado de *Cover Test Alternado* (CTA). Para a realização do CTA, o ocluidor é movido de um olho para o outro, executando dessa forma o processo de oclusão de ambos os olhos de maneira alternada. Diferente do *cover test* simples, os dois olhos são analisados durante uma oclusão. Quando o ocluidor está ocluindo o olho esquerdo, por exemplo, observa-se o comportamento do olho direito, contudo, ao remover o ocluidor é preciso atenção também ao comportamento do olho que está sendo desobstruído, pois ele pode apresentar movimento, que pode ser pequeno e muito rápido. Conforme já dito, o CTA é destinado para a detecção de desvios latentes, ou seja, que não são possíveis detectar apenas observando o paciente. Por fim, a magnitude e direção são medidas através da observação do movimento do olho.

Figura 6 – Execução do *Cover test* alternado.

Fonte: Adaptado de Faruk (2015).

3.4 Redes neurais artificiais

As redes neurais artificiais podem ser definidas como modelos matemáticos, abstraídos por conjuntos de unidades computacionais ligadas entre si, que realizam processamento simples e cujo funcionamento é inspirado nos neurônios biológicos. Esses modelos são

muito explorados para a realização de reconhecimento de padrões (LIMA et al., 2020; DONG; WANG; ABBAS, 2021; CRUZ et al., 2022).

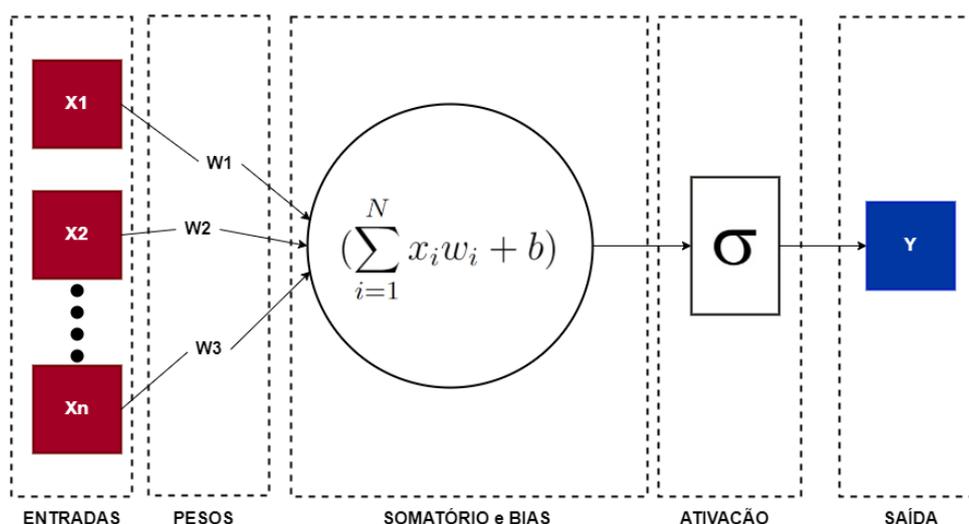
Esse campo da computação remete a descoberta das células nervosas do cérebro no século XIX, chamadas de neurônio. Mas apenas nas décadas de 20 a 40, já no século XX, foi possível compreender profundamente seu funcionamento, assim os primeiros modelos matemáticos descrevendo esse funcionamento foram propostos (KOVÁCS, 2002). Em 1943, McCulloch e Pitts (1943) publicaram um artigo que pode ser considerado pioneiro na abordagem de criação de estruturas artificiais. Essas estruturas, buscavam emular o funcionamento do neurônio biológico. Desta forma, propuseram um neurônio simples, conhecido como neurônio de McCulloch, que tratava-se de um dispositivo binário, o qual processava expressões booleanas. Tendo como base esses conhecimentos, em 1950 Rosenblatt (1958) na universidade de Cornell, propôs uma rede chamada *perceptron*, precursora do que hoje é conhecido como Redes Neurais Artificiais (RNA).

Na *perceptron*, a unidade básica de processamento de uma RNA, o neurônio artificial representa uma função matemática, na qual dado um conjunto de entrada é fornecida uma saída. Esse processo é descrito pela Equação 1

$$f(x) = \sigma\left(\sum_{i=1}^N x_i w_i + b\right) \quad (1)$$

onde x_i é a entrada i , w_i é o peso associado a entrada i , b é o termo *bias*¹ e o σ é a função de ativação. Outra forma de visualização da arquitetura da *perceptron*, vista em etapas, é mostrada na Figura 7.

Figura 7 – Etapas da arquitetura da *perceptron*.



Fonte: Próprio autor.

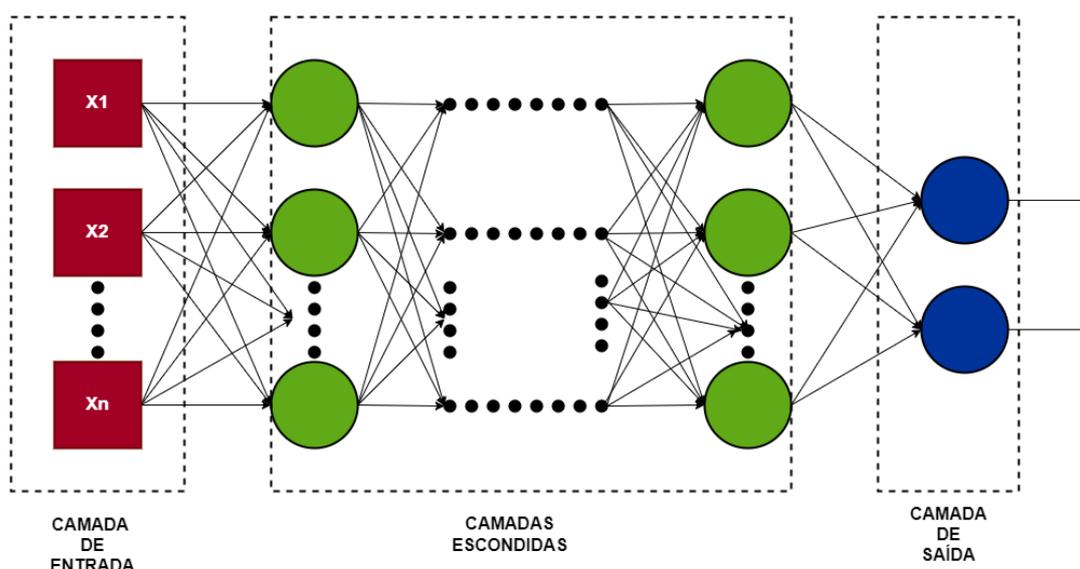
¹ O "bias" é um elemento que serve para aumentar o grau de liberdade dos ajustes dos pesos.

3.4.1 Multilayer Perceptron

Multilayer perceptron (MLP) é uma RNA que se difere da *perceptron* por possuir múltiplas camadas de neurônios em sua arquitetura. Essas camadas são completamente conectadas à camada seguinte. Por sua vez, os neurônios implementam funções de ativação. As redes que seguem essa estrutura podem aproximar qualquer função mensurável e para qualquer precisão que se deseje (HORNÍK; STINCHCOMBE; WHITE, 1990).

De modo geral, a MLP é definida conforme a Figura 8. Na camada de entrada são recebidos os dados de entrada do modelo. Logo em seguida, tem-se uma ou mais camadas intermediárias, também chamadas de camadas escondidas. Por fim, existe a camada de saída. Como dito anteriormente, cada camada é completamente conectada com a seguinte, mas não possui ligação com ela mesma, dessa forma é enquadrada na categoria conhecida como *feedforward*.

Figura 8 – Arquitetura básica *Multilayer perceptron*.



Fonte: Próprio autor.

Ao utilizar a rede para a resolução de um problema é necessário treiná-la. No treinamento de uma MLP, define-se uma função de erro afim de calcular o erro das predições realizadas com base em valores de saída esperados. Esse processo também é chamado de aprendizado supervisionado. O objetivo do treinamento é a minimização do erro total da rede, o processo responsável por esse comportamento é o algoritmo de retropropagação no inglês chamado de *Backpropagation*.

3.4.2 Backpropagation

A *backpropagation* possui duas grandes etapas: na primeira etapa, os dados são apresentados na camada de entrada e passam através de toda a rede sendo transformados

por cálculos, até chegarem a camada de saída; a segunda etapa consiste em comparar a saída obtida, com exemplos de saída esperada e calcula-se um valor de erro. Logo em seguida, esse erro é propagado no sentido inverso, da camada de saída para a camada de entrada, ajustando os pesos das conexões entre as camadas escondidas (RUMELHART; HINTON; WILLIAMS, 1986).

De forma mais detalhada, no início do algoritmo de retropropagação realiza-se o cálculo do erro de saída. Logo após, calcula-se o resultado do neurônio K na camada de saída da iteração $p(y_k(p))$, então calcula-se também um sinal de erro ($e_k(p)$) para este neurônio na iteração em questão, subtraindo o resultado esperado naquela iteração. Este processo é descrito por $y_{d,k}(p)$, conforme mostra a Equação 2.

$$e_k(p) = y_{d,k}(p) - y_k(p) \quad (2)$$

O sinal de erro calculado é propagado para trás através da extensão da rede, sentido inverso do *feedforward* explicado anteriormente. Esse processo possibilita um ajuste nos valores dos pesos presentes nas camadas, exemplificados pelas letras w na Figura 7. Vale ressaltar que o ajuste de pesos é realizado de forma diferente nas camadas de saída (camada k), ocultas (camadas j) e por fim a de entrada (camada i).

Para a camada de saída, calcula-se o gradiente do erro no neurônio k na iteração p . Além disso, usa-se os valores de ($e_k(p)$) e ($e_k(p)$), calcula-se a soma ponderada das entradas do neurônio k na iteração p , representado por $X_k(p)$, conforme a Equação 3. O valor $\delta_k(p)$ é o gradiente de erro para os neurônios na camada de saída. Com esses valores calculados, são feito os ajustes dos pesos a partir da camada de saída em direção a camada de entrada.

$$\delta_k(p) = \frac{\delta y_k(p)}{\delta X_k(p)} \cdot e_k(p) \quad (3)$$

O valor $\delta_k(p)$ é utilizado para calcular a correção dos pesos na iteração p e atualiza-se o valor dos pesos a serem usados na iteração seguinte ($p+1$), como é mostrado na Equação 4. A variável w_{jk} representa o peso entre o neurônio j da camada oculta, e o neurônio de k da camada de saída. Enquanto α representa a taxa de aprendizagem (parâmetro que pode ser fornecido pelo usuário) e y_j é a saída do neurônio j na camada interna.

$$W_{jk}(p+1) = W_{jk}(p) + \alpha \cdot y_j(p) \cdot \delta_k(p) \quad (4)$$

Em seguida, após atualizar os pesos dos neurônios que incidem na camada de saída, atualiza-se os pesos nas camadas internas. Para isso, calcula-se o gradiente de erro dos neurônios na camada escondida, conforme a Equação 5. O gradiente do erro do neurônio j

da camada ocultas na iteração p é representado por $\delta_k(p)$ e w_{jk} já foram calculadas nas Equações 3 e 4.

$$\delta_j(p) = y_j(p) \cdot [1 - y_j(p)] \cdot \sum_{k=1}^l \delta_k(p) w_{jk}(p) \quad (5)$$

Dado o gradiente, atualiza-se os pesos dos neurônios das camadas escondidas, através da Equação 6. O peso entre o neurônio i da camada de entrada e o neurônio j da camada oculta é representado por w_{ij} . Se seu valor é obtido de forma semelhante aos pesos da camada de saída. Contudo, na camada escondida, usa-se o termo x_i , que representa a soma ponderada das entradas do neurônio j , vinda dos neurônios da camada de entrada.

$$w_{ij}(p+1) = w_{ij}(p) + \alpha \cdot x_i(p) \delta_j(p) \quad (6)$$

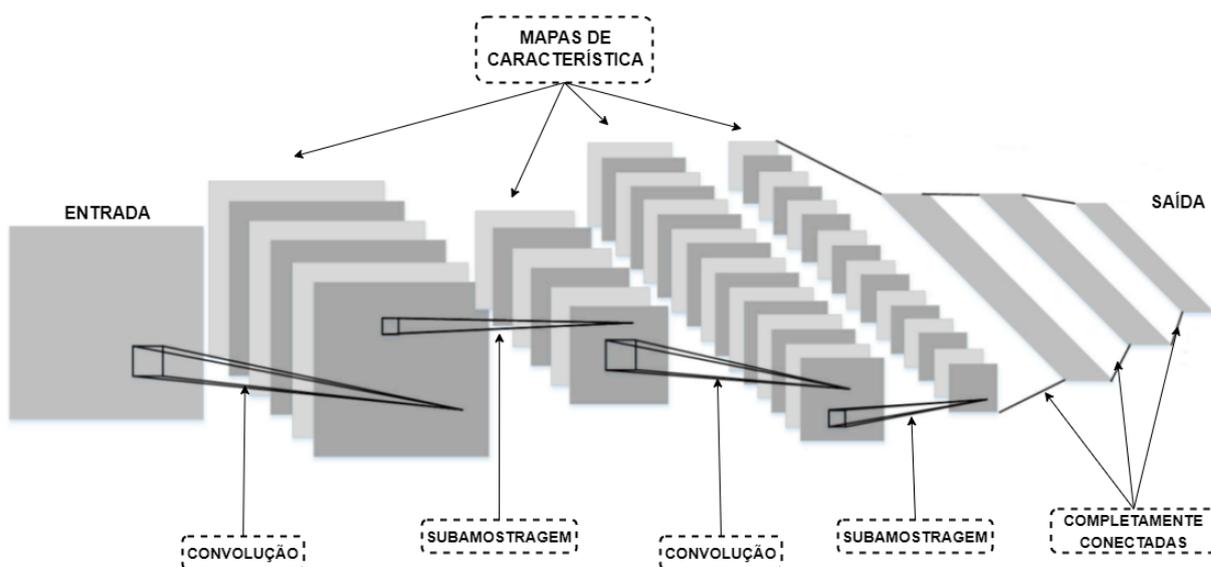
Por fim, após ser feita a atualização da camada escondida, repete-se o processo de *feedforward* seguido pelo *backpropagation* até que o parâmetro máximo de iterações, também chamado de épocas, seja atingido.

3.4.3 Redes neurais convolucionais

As redes neurais convolucionais, do inglês *Convolutional neural network* (CNN), são RNAs de múltiplas camadas que utilizam o aprendizado profundo para a solução de problemas. Surgiram pelo interesse em ter camadas ocultas mais profundas e superar o desempenho dos métodos clássicos em diferentes áreas, especialmente no reconhecimento de padrões. O aspecto mais benéfico das CNNs é reduzir o número de parâmetros. A redução da quantidade de parâmetros, alcançadas pelas CNNs, permitiu a pesquisadores construírem modelos para resolver tarefas mais complexas, o que não era possível com as RNAs clássicas. A suposição mais importante sobre os problemas resolvidos pela CNN, são que não deveria existir características que são espacialmente dependentes e permitem obter recursos abstratos quando a entrada se propaga em direção as camadas mais profundas (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

Uma CNN normalmente tem sua arquitetura baseada em três tipos de camadas, sendo elas: camada de convolução, subamostragem e camada completamente conectada (GUO et al., 2016). A Figura 9 mostra a organização dessa arquitetura. As camadas de convolução e subamostragem são as camadas no início da rede, são responsáveis pela geração dos chamados mapas de características. Quase no final da rede estão localizadas as camadas completamente conectadas, responsáveis pela classificação dos padrões. Por fim, encontra-se a saída da rede.

Figura 9 – Arquitetura base de uma CNN.



Fonte: Adaptado de Tra et al. (2017).

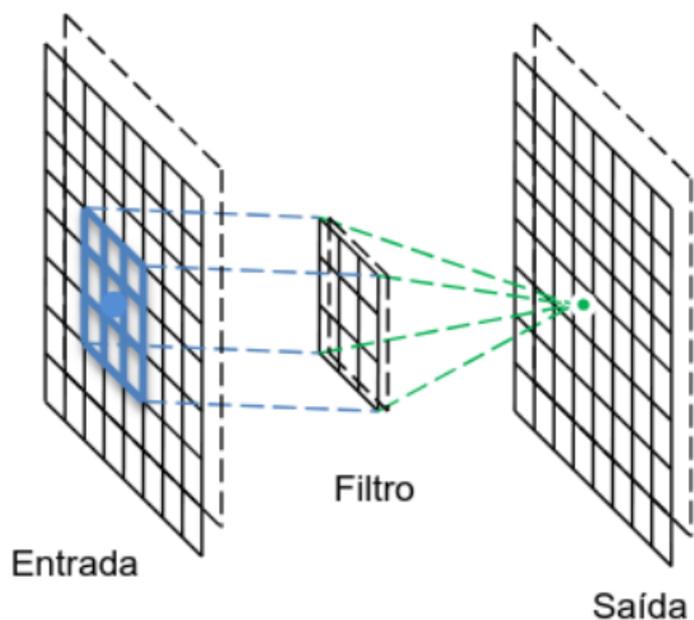
A convolução é um operador linear que, a partir de duas funções dadas, resulta em uma terceira que mede a soma do produto dessas funções. Essa medição, é realizada ao longo da região subentendida pela superposição delas em função do deslocamento existente entre elas (PEDRINI; SCHWARTZ, 2008). Basicamente trata-se de uma operação local entre pixels, em que uma matriz é tratada por outra chamada de *kernel*. A convolução é muito utilizada no processamento de imagens digitais como, por exemplo, os filtro de suavização e de realce.

No âmbito das CNNs, a camada de convolução possui filtros que podem ser treinados e aplicados sobre a imagem de entrada. Como resultado tem-se outras matrizes chamadas de mapas de características ou recursos (GUO et al., 2016). Durante o treinamento da rede cada filtro aprende como detectar uma característica específica, podendo ser da imagem como um todo ou de partes dela. O processo de convolução é apresentado na Figura 10.

De forma subsequente às camadas de convolução, existem as camadas de subamostragem. Essas camadas são utilizadas para diminuir as dimensões dos mapas de características gerados no processo de convolução. Seu funcionamento é similar ao processo executado na camada anterior. Para a subamostragem é realizado à aplicação de filtros sobre os mapas, extraindo amostras dessas regiões. Essa extração, pode utilizar a média aritmética (*Average Pooling*) ou o valor máximo do *kernel* (*Max Pooling*). O *Max pooling*, por exemplo, realiza a extração do valor máximo da janela de convolução, o processo de *Max pooling* está ilustrado na Figura 11.

No fim da CNN, encontra-se, como dito anteriormente, a camada completamente conectada. Essa camada é semelhante as camadas das MLPs. Nela, os neurônios de entrada recebem pixels dos mapas de recursos e geram classificação dos padrões. Neste trabalho é

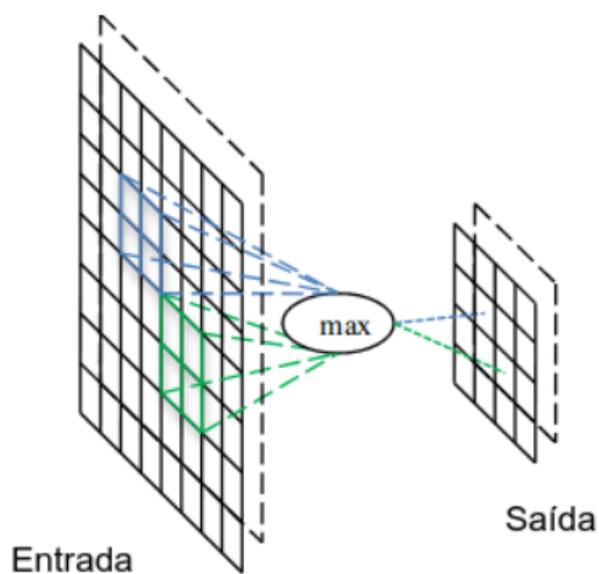
Figura 10 – Execução de uma convolução.



Fonte: Adaptado de Guo et al. (2016).

Figura 11 – Execução do *Max pooling*.

MAPA DE RECURSOS



Fonte: Adaptado de Guo et al. (2016).

utilizada uma arquitetura baseada nos conceitos das redes neurais convolucionais, chamada *You Only Look Once* (YOLO). A YOLO é empregada na etapa de detecção dos olhos e ocluser (Seção 4.2).

3.4.4 YOLO - *You Only Look Once*

You Only Look Once (Você só olha uma vez), abreviado através da sigla YOLO, é uma rede neural convolucional projetada para a detecção de objetos em tempo real proposta por Redmon et al. (2016). A detecção de objetos, seja em imagens estáticas ou em vídeos, consiste em apontar o local na imagem no qual o objeto desejado se encontra, além de classificá-los. Outras abordagens para detecção de objetos como, por exemplo, a proposta por Ren et al. (2015), detectam objetos baseados em regiões. Para esse tipo de detecção, primeiro é necessário gerar regiões nas quais é possível haver um objeto de interesse e posteriormente classificar essas regiões. Algoritmos como o *Selective search* (UIJLINGS et al., 2013) e *Support vector machine* (CORTES; VAPNIK, 1995) foram utilizados para essas tarefas, respectivamente. Os principais problemas dessas abordagens estavam ligados ao treinamento complexo e lento.

Diferente da abordagem anterior, a YOLO não trata a detecção de objetos como um problema de classificação, mas sim como um problema de regressão (REDMON et al., 2016). Através do modelo, a YOLO de forma unificada prevê regiões nas quais o objeto se encontra e em paralelo informa a classe que o objeto pertence. Dessa forma, a YOLO processa imagens em altas taxas de quadros por segundo (FPS - do inglês: *frames per second*) e sua versão reduzida alcança até 155 FPS.

A Figura 12 apresenta o fluxo de execução da YOLO. No início do processamento a imagem de entrada é dividida em uma grade de células $S \times S$. Para cada célula da grade a rede prevê um número B de *Bouding boxes* e a confiança (probabilidade) dessa região em conter um objeto. Caso a célula não abrigue nenhum objeto o valor da confiança deve ser zero. Caso contrário, se naquela célula existir um objeto, o valor da confiança deve refletir o quanto a *bounding box* encontrada representa a região de interesse. Esse valor de confiança é calculado por meio da Equação 7.

$$\text{Confiança} = Pr(\text{Object}) \times \text{IOU} \quad (7)$$

onde $Pr(\text{Object})$ é a probabilidade de existir um objeto na célula e a IOU é a interseção sobre a união da *bounding box* predita com a real. Se o valor de confiança for acima de um limiar estipulado, a rede criará um mapa de probabilidades de classes, para prever qual o tipo de objeto encontrado.

O mapa de probabilidade é calculado conforme a Equação 8. Em que, para cada célula são calculadas C probabilidades condicionais para cada classe no conjunto de dados.

$$C = Pr(\text{Classe}_i / \text{Objeto}) \quad (8)$$

onde i varia de 1 a quantidade de classes existentes.

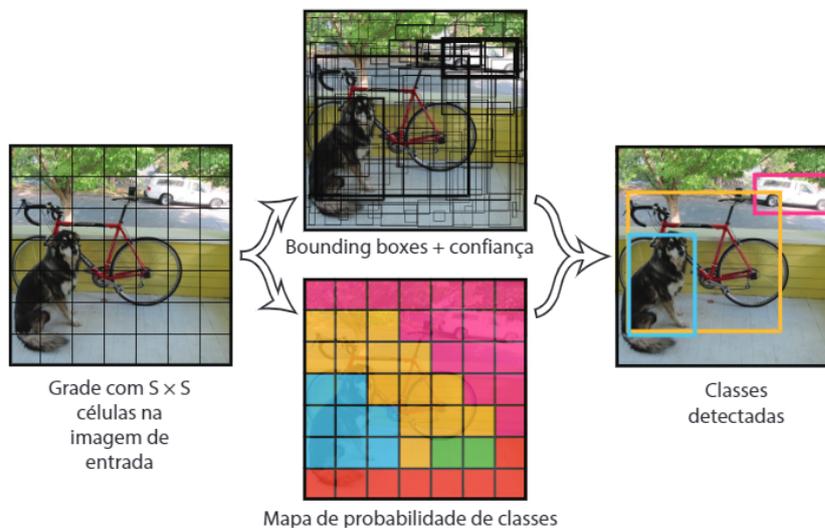
Logo em seguida, cada probabilidade C é multiplicada pelo valor de confiança para uma classe específica em uma *Bounding box*, esse processo é mostrado na Equação 9.

$$Pr(Classe_i * IOU) = Pr(Classe_i/Objeto) * Pr(Objeto) * IOU \quad (9)$$

Outro artifício utilizado pela YOLO para melhorar a detecção de objetos é o conceito de âncoras. As âncoras são retângulos com razões entre largura e altura pré-definidos e que representam as formas ideais dos objetos os quais deseja detectar (REDMON et al., 2016). Dessa forma, não é necessário prever diretamente o tamanho de uma *bouding box*, a rede apenas ajusta as âncoras de acordo com cada uma das células.

Para o desenvolvimento deste trabalho foi utilizada a quinta versão da YOLO, chamada de YOLOv5, a qual será explicada na próxima seção. Esta CNN foi escolhida por apresentar boa precisão para a detecção de múltiplos objetos, capacidade de detectar objetos pequenos, otimização do consumo de recursos e treinamento em espaços curtos de tempo.

Figura 12 – Pipeline de execução da YOLO.



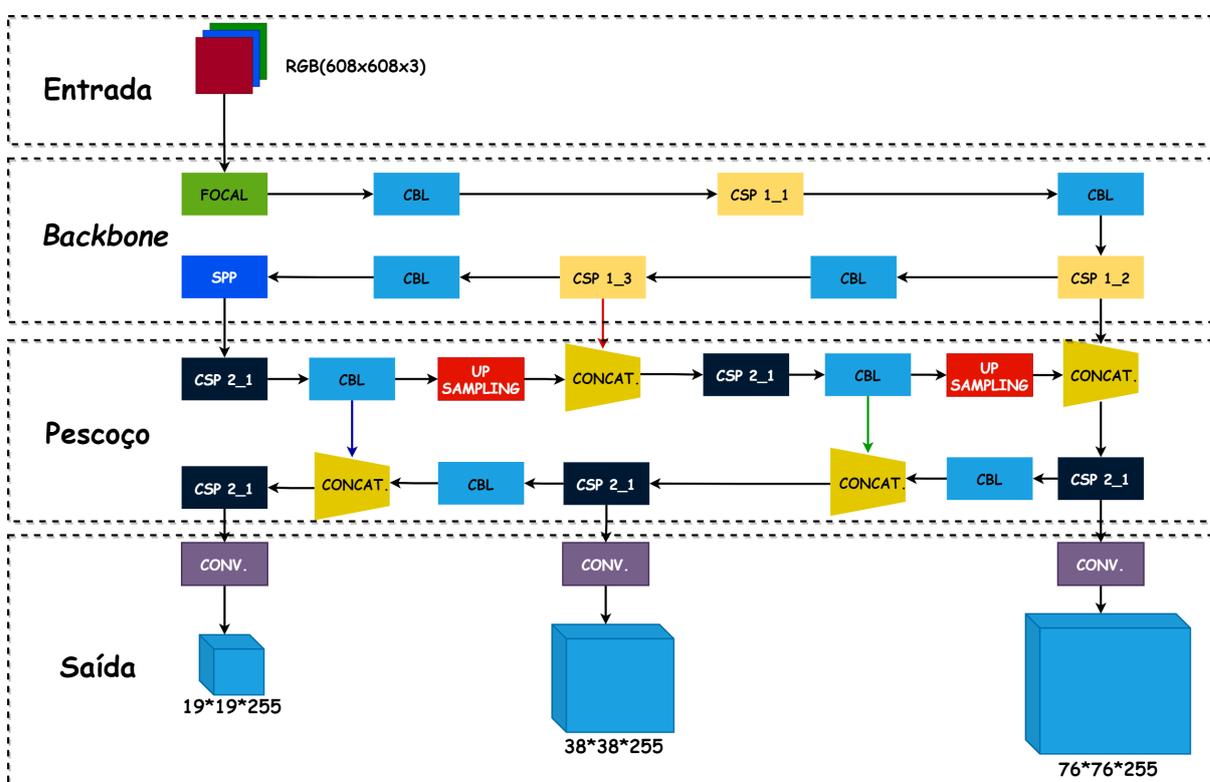
Fonte: Adaptado de Redmon et al. (2016).

3.4.5 YOLO Versão 5

A rede neural convolucional YOLO, como explicado na Seção 3.4.4, trouxe consigo uma nova abordagem para a detecção de objetos e introduziu uma nova família de CNNs. Contudo, por se tratar de uma inovação, havia espaços para exploração do conceito e melhorias. Desta forma, outras arquiteturas - também chamadas de gerações - da YOLO foram propostas. O trabalho proposto por Redmon et al. (2016) se tornou a primeira geração, de forma subsequente vieram a segunda geração (YOLOv2) (REDMON; FARHADI, 2017), terceira geração (YOLOv3) (REDMON; FARHADI, 2018), quarta geração (YOLOv4) (BOCHKOVSKIY; WANG; LIAO, 2020) e por fim, encontra-se na sua quinta geração (YOLOv5) (JOCHER et al., 2021).

Quando se analisa as arquiteturas da YOLOv5 percebe-se semelhanças com sua predecessora, mas existem diferenças. A arquitetura da YOLOv5 pode ser dividida em quatro grandes partes, são elas: Entrada da rede, *Backbone*, *Pescoço* (*Neck*) e Saída (camada de previsão). A organização dessas etapas e seus relacionamentos são mostrados na Figura 13 e serão explicados de forma individual nas subseções seguintes.

Figura 13 – Arquitetura da YOLOv5.



Fonte: Próprio autor.

3.4.5.1 Camada de entrada

A primeira grande etapa da arquitetura da YOLOv5 é a de entrada. Essa etapa da rede recebe os dados de entrada e realiza os primeiros processamentos. Na entrada da rede são utilizadas três técnicas, sendo elas: aprimoramento de dados em mosaico, Cálculo de âncoras adaptativos e dimensionamento adaptável de imagem.

O aprimoramento de dados em mosaico, foi proposto por Glenn Jocher como contribuição para o repositório original da YOLO. Essa técnica foi utilizada pela primeira vez no trabalho publicado por [Bochkovskiy, Wang e Liao \(2020\)](#) e utiliza seleção de hiperparâmetros através de algoritmo genético para resolver o problema de *grid sensitivity*. Essa técnica é ideal quando o conjunto de dados, possui muitos alvos pequenos e uma distribuição desigual desses alvos. Dentre as vantagens da aplicação do aprimoramento de dados em mosaico, destaca-se o enriquecimento do conjunto de dados, através do uso aleatório de imagens, dimensionamento aleatório e distribuição aleatória de objetos ([ADIBHATLA et al., 2021](#)). A etapa de dimensionamento aleatório adiciona muitos alvos pequenos, tornando a rede mais robusta.

No algoritmo YOLO, descrito na Seção 3.4.4, existe o conceito de âncoras. As âncoras são retângulos com razões entre largura e altura pré-definidos, utilizados para representar formas de objetos. Nas primeiras versões da rede foram utilizadas âncoras pré-definidas e de forma fixa. No entanto, para diferentes conjuntos de dados haverá imagens e objetos com tamanhos variados, que podem não ser bem representados por âncoras pré-definidas. Para solucionar esse problema, a YOLOv5 utiliza uma função incorporada no código, a qual se aproveita das marcações das imagens de treinamento (*Ground Truth*), para a geração de âncoras adaptadas, para os objetos presentes na base de dados para o determinado problema. Além disso, essas âncoras são atualizadas durante a etapa de treinamento.

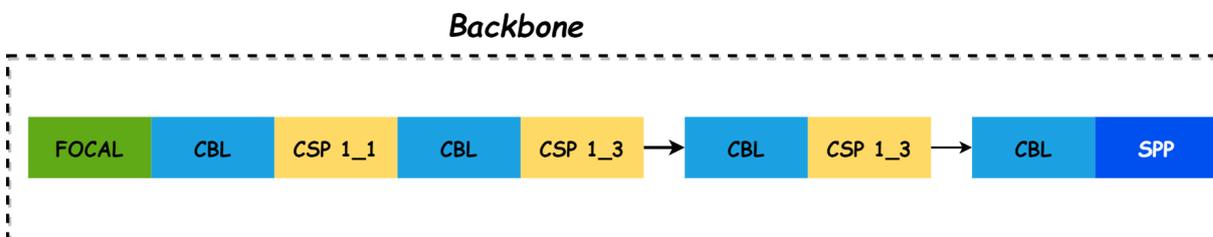
O último recurso da camada de entrada, é o dimensionamento adaptável de imagem. Na detecção de objetos utilizam-se imagens, que na maioria das vezes não possuem comprimentos e larguras iguais. Portanto, é comum a realização do redimensionamento de imagens para a inserção posterior na rede. Muitos dos algoritmos utilizados para redimensionamento causam distorções na imagem resultante, podendo afetar a qualidade, *aspect ratio*² e até mesmo seu conteúdo em casos que são inseridos pixels pretos para preenchimento.

² Relação de proporção entre a largura e altura da imagem.

3.4.5.2 Backbone

Logo após a etapa de entrada da YOLOv5, encontra-se o *Backbone*. Nessa fase da arquitetura destacam-se a estrutura de foco e o uso da *Cross Stage Partial network* (CSP) e é estruturado conforme exibido na Figura 14. A primeira camada é chamada de focal, de forma subsequente encontra-se uma camada composta por um conjunto de operações (*convolution, batch normalization e Leaky ReLU*) abreviada por (CBL), logo após uma estrutura CSP e no fim do *backbone* observa-se uma camada *Spatial Pyramid Pooling* (SPP). Mais detalhes dos blocos que compõem o *backbone* serão explicados no decorrer desta seção.

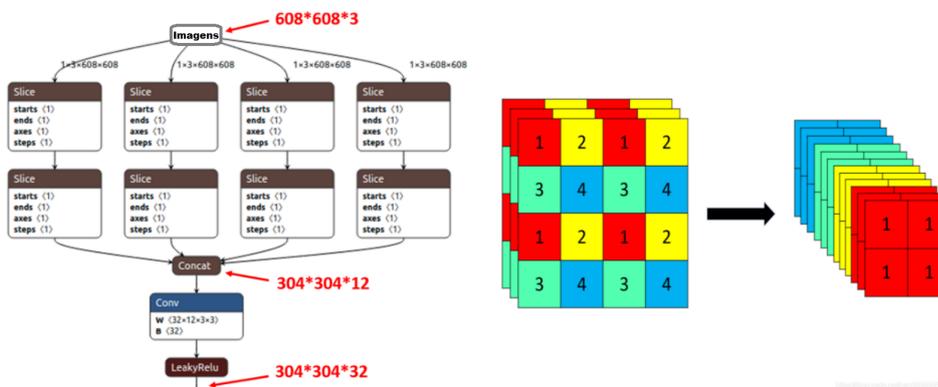
Figura 14 – Diagrama do *backbone* YOLOv5.



Fonte: Próprio autor.

A camada de foco, segundo (JOCHER et al., 2021), criador da YOLOv5, tem como objetivo a substituição das três primeiras camadas da YOLOv3. Essa substituição, ainda segundo ele, reduz camadas, parâmetros, memória CUDA utilizada, aumenta a velocidade de avanço e retrocesso, e impacta minimamente a qualidade dos resultados. A camada de foco é detalhada na Figura 15.

Figura 15 – Camada de foco.



Fonte: Adaptado de Adibhatla et al. (2021).

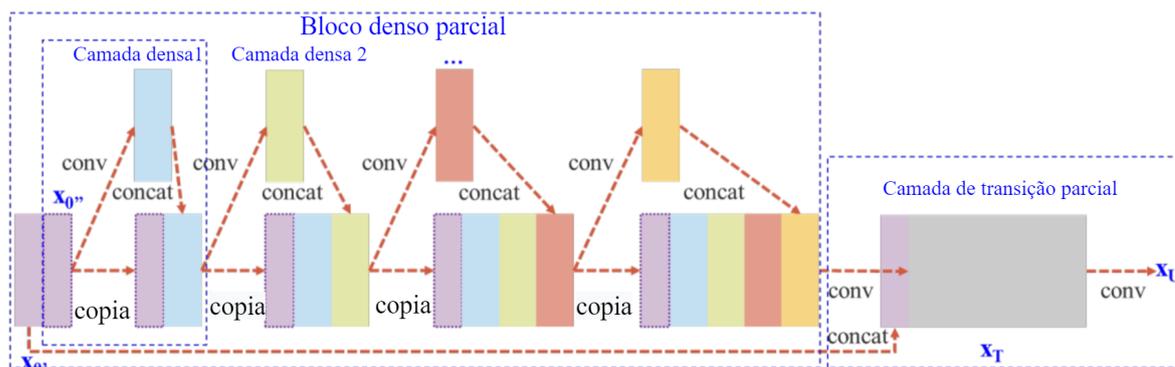
Na Figura 15 é possível observar que, por exemplo, uma imagem com dimensões 608 x 608 x 3 é alimentada na estrutura de foco, então a operação de fatiamento é usada para alterar a imagem para um mapa de recursos com dimensões 304 x 304 x 12, além de 32 *kernels* para serem utilizados nas operações de convolução.

3.4.5.3 Estrutura CSP - CSP Network

Outro recurso importante empregado no *backbone* da YOLOv5 é a CSP Net (*Cross Stage Partial Network*). Segundo Wang et al. (2020), a rede é apresentada como uma nova solução para mitigar o problema da exigência de cálculos de inferência pesados. Segundo o autor, o problema está ligado as informações duplicadas no gradiente de otimização das redes anteriores. Como resultado a CSP Net apresenta ganhos de até 20% quando comparada a outras abordagens.

O objetivo no projeto da CSP Net é permitir que a arquitetura obtenha uma combinação de gradiente mais rica, enquanto paralelamente é reduzida a quantidade de cálculos. Para isso, o mapa de recursos da camada base é particionado em duas partes, e uma delas é concatenada diretamente à camada de transição parcial. Esse processo é mostrado na Figura 16.

Figura 16 – Bloco residual parcial da CSP Net



Fonte: Adaptado de Wang et al. (2020).

No *Backbone* da YOLOv5 a primeira camada do bloco chama-se CBL (Figura 17), trata-se do menor módulo da arquitetura principal da YOLO. Esse módulo é composto por uma camada de convolução, seguida por uma de *batch normalization* e por fim por uma ativação *Leaky ReLU*.

Figura 17 – Estrutura do bloco CBL.



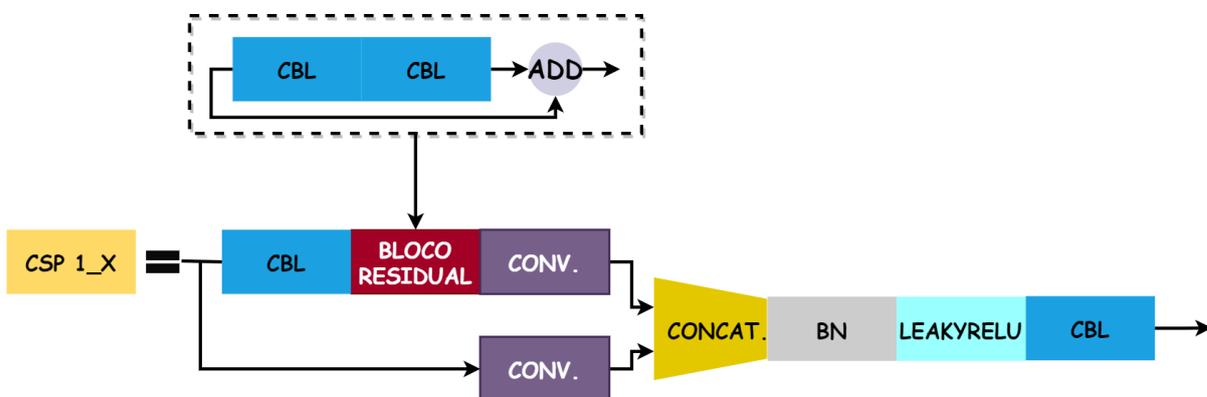
Fonte: Próprio autor.

Logo após o primeiro bloco CBL, encontra-se o módulo CSP1_X que representa a CSP Net e é detalhado na Figura 18. Esse módulo contém blocos CBL e unidades residuais³ e objetiva extrair melhor as características profundas da imagem. Vale ressaltar,

³ Cada unidade residual é composta de dois módulos CBL e sua saída é a soma da saída dos dois módulos CBL e a entrada original.

que o valor de X no rótulo do bloco representa a quantidade de unidades residuais.

Figura 18 – Bloco CSP1_X.

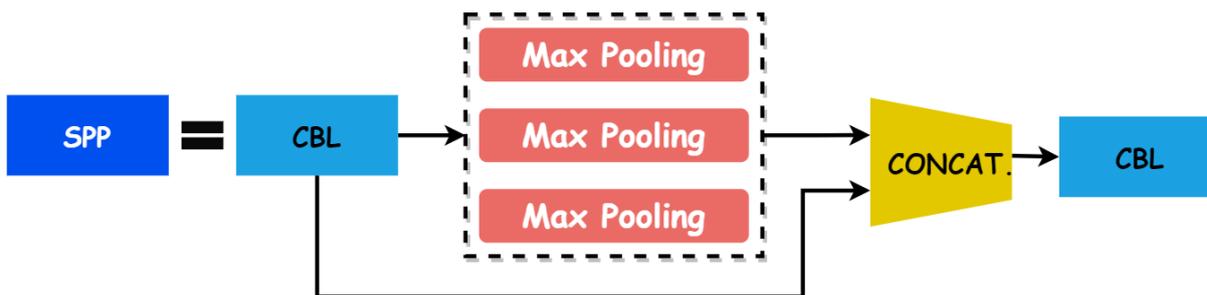


Fonte: Próprio autor.

O fluxo do dado no bloco CSP1_X é a seguinte: em primeiro lugar, a entrada inicial é inserida em uma bifurcação; no primeiro ramo a entrada é transformada por um bloco CBL, bloco residual e uma convolução; no outro ramo a entrada sofre apenas uma convolução. O resultado das duas convoluções são concatenados. A saída da concatenação então é normalizada, ativada com a função *Leaky ReLU* e por fim passa por um último bloco CBL. Importante ressaltar que o mapa de saída tem as mesmas dimensões do mapa de recursos da entrada.

A última camada que compõem o *backbone* YOLOv5 é representada na Figura 19. Essa camada é nomeada de *Spatial Pyramid Pooling* (SPP). O objetivo desta camada é transformar o mapa de características, que possui apenas uma resolução, em um vetor com mesmas dimensões da camada completamente conectada. Nessa estrutura o fluxo é da seguinte forma: primeiramente, o mapa de recursos é processado por um bloco CBL; em sequência, a operação de *max pooling* é executada em paralelo através de três camadas; na terceira etapa, o mapa de recurso da primeira etapa é concatenado com o resultado da segunda etapa; na última fase, o resultado da concatenação é processado novamente por um bloco CBL.

Figura 19 – Estrutura da *Spatial Pyramid Pooling*.

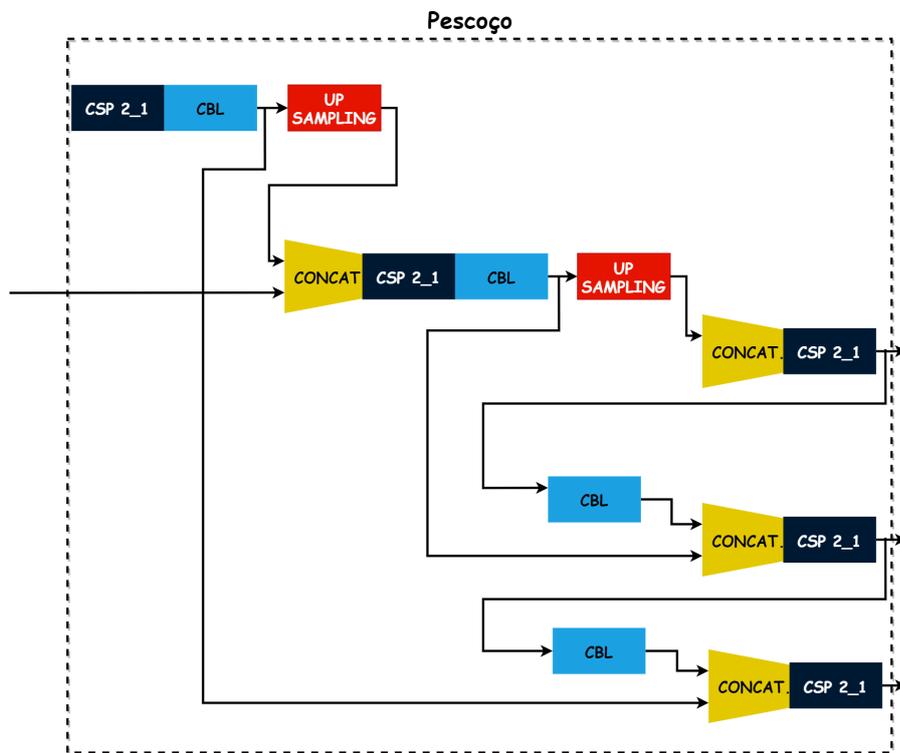


Fonte: Próprio autor.

3.4.5.4 Neck - Pescoço

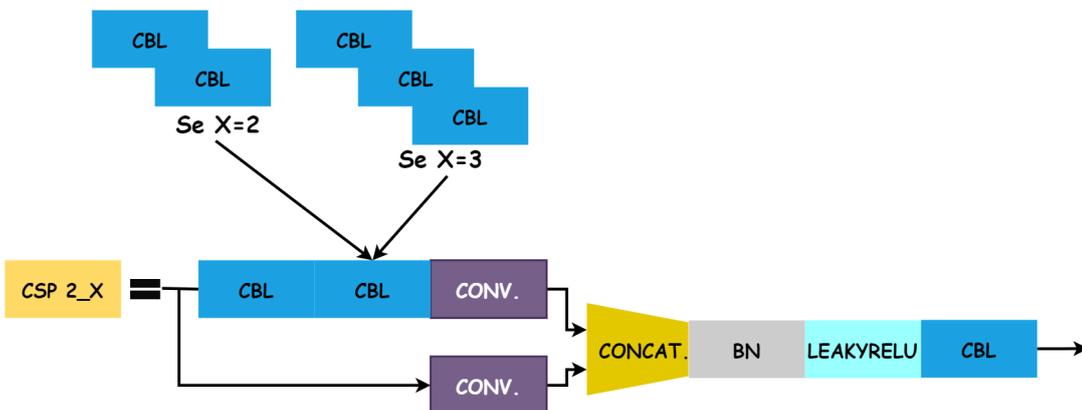
A arquitetura da etapa que compreende o pescoço da YOLOv5 é apresentada na Figura 20. Nessa estrutura a rede também utiliza primariamente blocos CSP. Contudo, esse bloco chamado de CSP2_X apresenta uma importante diferença em relação ao bloco CSP1_X. A diferença é o significado do valor de X. Enquanto, que no *backbone* X representa a quantidade de blocos residuais, no *neck* ele representa a quantidade de blocos CBL, conforme ilustrado na Figura 21.

Figura 20 – Estrutura do Neck da YOLOv5.



Fonte: Próprio autor.

Figura 21 – Bloco CSP2_X.



Fonte: Próprio autor.

Outro detalhe importante nessa parte da arquitetura, é a presença de operações de *upsampling*. Antes dos mapas de recursos passarem pelo primeiro bloco de *upsampling* são concatenados diretamente com resultados do bloco CBL, permitindo a detecção de objetos pequenos. Antes da execução do segundo bloco de *upsampling*, os mapas da entrada são concatenados diretamente com bloco CBL para preservar características de média resolução. Por fim, os mapas que passaram por dois processos de *upsampling*, são utilizados para a detecção de objetos em alta resolução.

3.4.5.5 Camada de saída

Na camada de saída da rede, serão realizadas as previsões das *bounding boxes*. Nessa camada, a YOLOv5 utiliza duas técnicas principais, são elas: A função de perda GIoU e a Supressão de não máximos. A supressão de não máximos, basicamente descarta resultados, os quais o grau de confiança estão abaixo de um certo limiar.

Para entendimento geral da função de perda GIoU, primeiro é necessário uma breve explicação sobre a comparação de *bounding boxes*. A interseção sobre a união, ou *Intersection over Union*(IoU), é utilizada na detecção de objetos como função de perda. Essa função mede a semelhança entre a caixa delimitadora predita e a caixa delimitadora real. O IoU pode ser expresso através da Equação 10.

$$IoU = \frac{area(caixa(predita) \cap caixa(verdadeira))}{area(caixa(predita) \cup caixa(verdadeira))} \quad (10)$$

Onde, o valor do IoU varia entre 0 e 1.

Contudo, quando ocorre a situação em que caixas delimitadoras não se sobrepõem ou existem maneiras diferentes de sobreposição, ou seja, as parte sobrepostas tem áreas equivalentes - mas a direção e conteúdo não são iguais - o IoU não é tão preciso. Portanto, a YOLOv5 utiliza uma variação do IoU. Essa variação é chamada de GIoU, *Generalized Intersection over Union*. O GIoU é descrito segundo a equação 11.

$$GIoU = \frac{|A \cap B|}{|A \cup B|} - \frac{|C \setminus (A \cup B)|}{|C|} = IoU - \frac{|C \setminus (A \cup B)|}{|C|} \quad (11)$$

Onde, A e B são duas caixas delimitadoras e C é o menor retângulo que pode conter A e B . O valor do GIoU, por sua vez varia entre -1 e 1.

3.5 Métricas de avaliação

A avaliação do desempenho de um método objetiva promover a validação dos resultados obtidos. Esta seção apresenta as métricas utilizadas para quantificar o desempenho

do método proposto. A avaliação dos resultados de classificação dos pacientes em saudável ou estrábico, a detecção do estrabismo, além de avaliar a magnitude e tipo de desvio, ou seja, o diagnóstico realizado. Considera-se, portanto, quatro situações possíveis.

- Verdadeiro Positivo (VP): O paciente é estrábico e o diagnóstico foi correto.
- Falso Positivo (FP): O paciente é estrábico e o diagnóstico foi incorreto.
- Verdadeiro negativo (VN): O paciente é saudável e o diagnóstico foi correto.
- Falso Negativo (FN): O paciente é saudável e o diagnóstico foi incorreto.

Neste trabalho, são utilizadas as medidas de Especificidade (E), Sensibilidade (S), Acurácia (AC) e Precisão (Pr) por serem métodos estatísticos comumente utilizados para validar resultados em problemas de classificação.

A sensibilidade (12) é a proporção de imagens corretamente classificadas como patológicas. A especificidade (13) indica a probabilidade de as imagens serem classificadas corretamente como saudáveis.

$$S = \frac{VP}{VP + FN} \quad (12)$$

$$E = \frac{VN}{VN + FP} \quad (13)$$

A acurácia (14) é a proporção total das amostras corretamente classificadas

$$AC = \frac{VP + VN}{VP + FP + FN + VN} \quad (14)$$

Além das métricas já citadas, também foi utilizado o *Mean Absolute Error* (MAE). O MAE é uma métrica que avalia o erro absoluto do valor estimado para o valor predito. No caso deste trabalho, empregou-se o MAE para a verificação do erro entre o valor calculado do desvio de cada paciente e o valor medido pelo especialista. O MAE é descrito calculado conforme a Equação 15.

$$MAE = \frac{1}{n} \sum_{t=1}^n |P_t - A_t| \quad (15)$$

onde P_t é o valor predito e A_t é o valor real.

3.6 Testes de hipóteses

Através da observação de dados em um determinado experimento, podem surgir as hipóteses. Essas hipóteses podem ser rejeitadas ou aceitas a partir de algum fator determinado. Portanto, esses testes são procedimentos estatísticos que permitem a tomada de decisão sobre duas ou mais hipóteses levantadas.

Neste trabalho, os testes de hipótese realizados utilizam o método de "teste Z" (LEHMANN; ROMANO, 2022). A partir desse teste, é possível determinar se a diferença entre duas proporções é significativa. Essas proporções, no contexto do trabalho, refere-se a medição do desvio ocular do paciente realizado pelo método proposto. Dessa forma, a abordagem consistem em quatro etapas: declarar as hipóteses, formular um plano de análise, analisar os dados da amostra e interpretar os resultados.

Durante a declaração das hipóteses são feitas afirmações sobre a diferença entre as proporções analisadas. Nesse caso, surge a hipótese nula, que deve ser aceita ou descartada. Em contrapartida, existe a hipótese alternativa, que é decorrente da rejeição da hipótese nula.

A formulação do plano de análise serve para descrever como utilizar os dados da amostra para rejeitar ou aceitar a hipótese nula. Nessa etapa, é definido o nível de significância α , que pode assumir valores entre 0 e 1. Por convenção, o nível recomendado é $\alpha = 0,05$. O teste Z utiliza os dados da amostra para calcular a proporção da amostra combinada (PAC) (Equação 16), o erro padrão (EP) (Equação 17), o escore z (EZ) (Equação 18) e o valor p (p) (Equação 19).

$$PAC = \frac{P1 \times N1 + P2 \times N2}{N1 + N2} \quad (16)$$

$$EP = \sqrt{PAC \times (1 - PAC) \times \left(\frac{1}{N1} + \frac{1}{N2}\right)} \quad (17)$$

$$EZ = \frac{P1 - P2}{EP} \quad (18)$$

$$p = 2 \times (1 - Norm.Dist(EZ)) \quad (19)$$

onde *Norm.Dist* é a distribuição normal.

Por fim, o valor p e o nível α são utilizados na etapa de interpretação dos resultados. Portanto, se $p > \alpha$, a hipótese nula é aceita, indicando que não há uma diferença significativa entre as proporções analisadas. Caso contrário, essa hipótese é rejeitada, levando a aceitação da hipótese alternativa.

3.7 Considerações

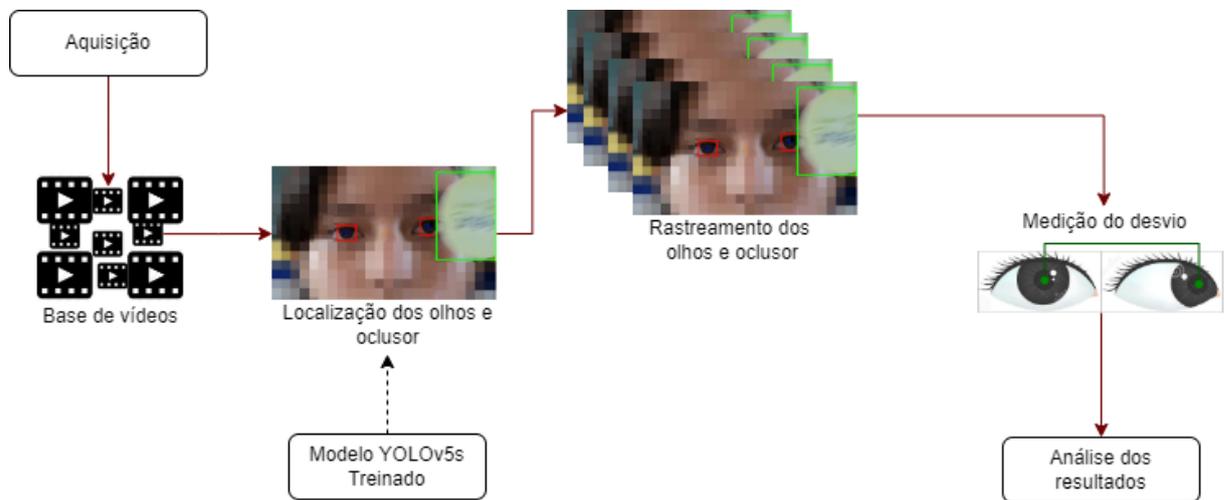
Nesse capítulo apresentou-se os fundamentos, conceitos e técnicas que fundamentam o desenvolvimento deste trabalho. Explicou-se as estruturas do olho humano, a natureza e classificações do estrabismo, o exame utilizado para diagnóstico - que é uma das bases utilizadas. Também foram abordados conceitos sobre redes neurais artificiais e uma de suas classes específicas, chamadas de redes neurais convolucionais. Por fim, foi explicada a CNN que originou uma nova família de redes neurais convolucionais e logo em seguida foi descrita a rede fundamental para o desenvolvimento deste trabalho, a YOLOv5.

No capítulo subsequente, é explicado como e onde cada conhecimento exposto no capítulo atual é empregado no método proposto, bem como os materiais utilizados no processo de construção do método que auxilia o diagnóstico do estrabismo através de vídeos do exame *cover test* alternado.

4 MATERIAIS E MÉTODO

Este capítulo descreve o método proposto para o diagnóstico automático do estrabismo por meio do exame *Cover test* alternado utilizando a rede neural YOLOv5. O método foi organizado em etapas. A primeira etapa refere-se à criação da base de vídeos (Seção 4.1), desde a aquisição dos vídeos até a análise e organização dos mesmos. Na segunda, realiza-se a detecção dos olhos e ocluser no vídeo do exame com o emprego da YOLOv5 (Seção 4.2). Por último, é realizada a medição do desvio do paciente (Seção 4.3). Um esquema resumido do método é apresentado na Figura 22.

Figura 22 – Visão geral do método proposto.



Fonte: Próprio autor.

4.1 Bases de vídeos

O método proposto foi avaliado em dois conjuntos de vídeos do exame *cover test* diferentes. A primeira base de vídeos foi a criada por Valente et al. (2017) em seu trabalho. Durante a produção deste trabalho, foi construída uma segunda base de vídeos. Uma terceira base de vídeos construída por Costa et al. (2019), foi utilizada para avaliar as etapas de treinamento da rede YOLO e a etapa de localização dos olhos.

4.1.1 Dataset de videos cover test I (CV-I)

A base utilizada no trabalho de Valente et al. (2017) é composta por vídeos do exame *cover test* de pacientes examinados em um consultório médico do Hospital da Universidade Federal do Maranhão (HU-UFMA), localizado na cidade de São Luís - MA,

Brasil. Para a construção da base de vídeos um protocolo para a obtenção dos vídeos foi elaborado. O protocolo estabelece uma distância de 50 cm entre o paciente e a câmera. Além disso, o paciente teve a cabeça apoiada em um suporte. Este suporte possui dois pontos de encaixe, um no queixo e outro na testa do paciente. Este suporte é projetado para evitar que a cabeça do paciente se desloque durante o teste. Por fim, a visão do paciente deve permanecer fixada na câmera e seu rosto deve estar centralizado. A câmera deve ser configurada para vídeos com resolução de 1920x1080 pixels com proporção de 16:9. Esta base de dados inclui 13 vídeos de pacientes que têm exclusivamente estrabismo horizontal.

4.1.2 Dataset de videos cover test II (CV-II)

A base de vídeos construída e utilizada neste trabalho é composta por vídeos de exames adquiridos em duas escolas públicas localizadas no município de São Luís - MA, Brasil. Os vídeos são de crianças, as quais a idade varia entre 10 e 15 anos, de ambos os sexos e de várias etnias. Para tal, para cada participante foi entregue um documento para ser assinado pelos pais ou responsáveis permitindo a utilização das imagens para fins de pesquisa. Este projeto foi aprovado pelo Comitê de Ética em Pesquisa do HU-UFMA e complementado pelo Certificado de Apresentação de Apreciação Ética, sob o número: 1.242.814. Para realizar de fato a aquisição dos vídeos, foi utilizada uma câmera digital Samsung Galaxy 2 com uma resolução de 16 *megapixels* e sistema operacional android. Durante a aquisição dos vídeos a câmera foi posicionada a uma distância aproximada de 50 cm, configurada para uma resolução de 1920 x 1080 *pixels* e com um aspecto de 16:9. Por se tratar de um exame oftalmológico é preciso que o paciente esteja com o olhar fixo para a câmera enquanto realiza o exame.

Para amenizar os movimentos do rosto, foi solicitado para que os pacientes apoiassem os cotovelos sobre a mesa, e apoiassem o queixo sobre os punhos cerrados. Ao todo foram obtidos 123 vídeos, os quais passaram por uma análise de qualidade. Foram verificados casos em que o paciente piscou de forma excessiva, fechou os olhos por um período longo, desviou o olhar para longe da posição da câmera e etc. Após essa avaliação os vídeos foram separados em dois grupos: o primeiro com 57 vídeos, que apresentam o mínimo de problemas já mencionados, os quais foram utilizados nos experimentos do método proposto e o segundo grupo com 66 vídeos, nos quais houve algum dos problemas citados no momento de aquisição, mas que foram úteis em etapas seguintes, como o treinamento da rede neural responsável pela detecção dos olhos e ocluser.

Quando se compara esse processo de aquisição de vídeos com o apresentado por [Valente et al. \(2017\)](#), destaca-se a não utilização de qualquer outro equipamento além da câmera. Não sendo necessária a utilização de equipamentos para fixação da cabeça

do paciente ou de iluminação, essa flexibilidade na aquisição dos vídeos pode facilitar a utilização do método proposto para a realização de triagens fora do consultório e em condições menos controladas.

4.1.3 Dataset de vídeos paralisia sexto nervo (PSN)

Com a finalidade de verificar a assertividade do modelo da YOLOv5 treinado. Foi utilizada a base de vídeos construída no trabalho desenvolvido por [Costa et al. \(2019\)](#) para o diagnóstico da paralisia do sexto nervo óptico. Essa base de dados é composta por 35 vídeos de voluntários, sendo 13 deles pessoas afetadas pela patologia e 22 vídeos são de pessoas saudáveis. Os pacientes que permitiram a gravação dos vídeos, assinaram um termo de consentimento. Este projeto tem autorização do comitê de ética em pesquisa do Hospital Universitário Federal do Maranhão através do parecer N^o 331/09, que garante o uso de imagens de humanos para fins de pesquisa acadêmica. Vale ressaltar que alguns dos vídeos foram adquiridos sem nenhum protocolo e com o uso de *smartphone*. Alguns outros, foram obtidos utilizando um protocolo definido. Este protocolo define que os vídeos podem ser adquiridos por uma câmera ou um celular apoiados em um tripé e os participantes apoiam o queixo sobre os punhos cerrados para minimizar o ruído do movimentos dos olhos.

4.2 Localização e rastreamento dos olhos e ocluser

O diagnóstico do estrabismo realizado pelo método proposto se baseia no cálculo da diferença de posições do centro da pupila durante o decorrer do exame, esse processo será explicado mais detalhadamente adiante. Essa forma de medição também é abordada de forma semelhante por outros trabalhos já mencionados na Seção 2.3.

Contudo, um dos principais problemas encontrados nessa abordagem é a dependência de uma boa localização do centro do limbo. A exemplo do trabalho apresentado por [Valente et al. \(2017\)](#), o qual dedica uma etapa de seu método apenas para a localização do limbo, uma etapa extensa e muito parametrizada, que acaba dependente de vários fatores externos como, por exemplo, iluminação do ambiente e um protocolo mais rígido para aquisição dos vídeos. Outros trabalhos (([WEBER et al., 2017](#)),([ZHENG et al., 2018](#)),([DERICIOĞLU; ÇERMAN, 2019](#)),([MIAO et al., 2020](#)), ([MEHRINGER et al., 2020](#))) também mostrados na Seção 2.3, buscam minimizar a dificuldade desta etapa através de marcações manuais, construção de hardwares específicos ou até mesmo óculos de realidade virtual, o que permite um melhor controle do cenário e conseqüentemente uma determinação dessa região de forma precisa, porém acabam encarecendo o método e dificultando a acessibilidade do mesmo.

Considerando a importância da etapa de localização da região que é englobada pelo limbo e buscando um método preciso mas que não precise de hardwares específicos, optou-se pelo emprego de uma rede neural convolucional de detecção de objetos para essa tarefa. Além disso, percebeu-se que além dessa região dos olhos também seria relevante o rastreamento do oclutor, ferramenta utilizada para ocluir o olho durante a execução do *Cover test*, portanto a rede escolhida precisa realizar uma detecção precisa, além disso, detectar múltiplos objetos em cena. Por esses motivos, optou-se pelo uso da rede neural convolucional para detecção de múltiplos objetos YOLO, considerada na literatura o estado da arte para esse tipo de tarefa, além de ter uma boa capacidade de rastreamento de objetos pequenos. Para o desenvolvimento do método foi empregada especificamente sua versão mais recente, a YOLOv5.

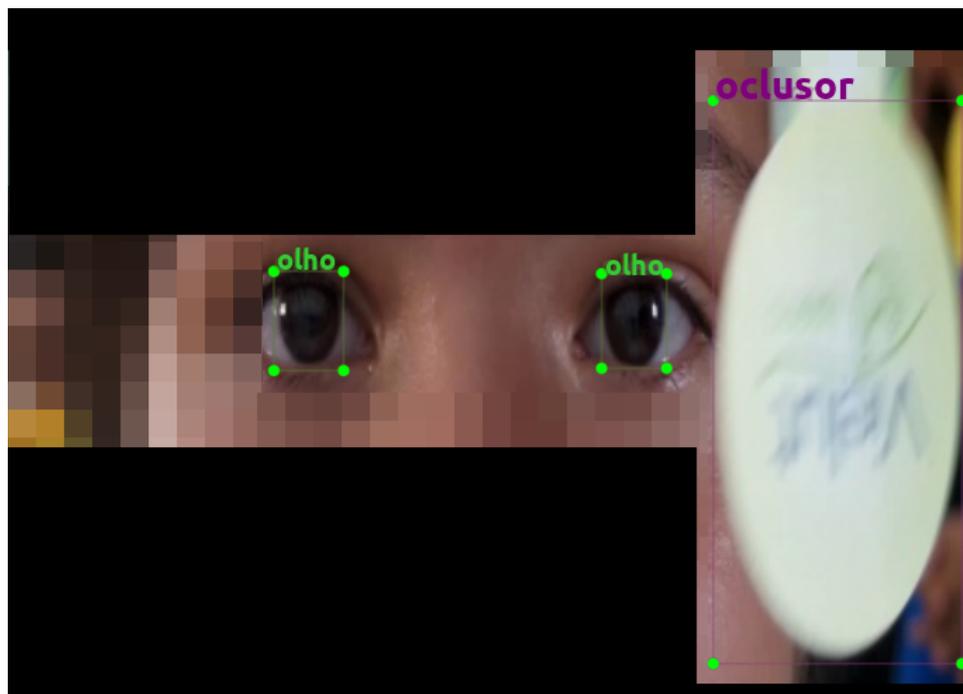
4.2.1 Treinamento da rede YOLOv5

Conforme explicado na Seção 4.1 alguns vídeos apresentaram problemas durante a execução do exame, por exemplo, vídeos interrompidos antes do término do exame, desvio da direção do olhar para longe do posicionamento da câmera, paciente se manteve de olhos fechados por um período longo, dentre outros fatores. Esses vídeos foram utilizados para compor a base de treinamento da rede neural utilizada na etapa de localização dos olhos. Para tal tarefa, foram extraídas imagens de cada *frame* do vídeo. Cada imagem foi submetida a uma probabilidade de 0,5 de ser rotacionada, espelhada ou ser aplicada sobre ela um *zoom*. Depois do aumento de dados, a base passou a ter 12.253 imagens. Dessas imagens, 70% delas (8.577) foram usadas para treino, 20% (2.450) para validação e 10% (1.226) para teste. As imagens foram separadas por paciente, para que não houvesse imagens do mesmo paciente no treino, validação e teste. Para cada imagem foram definidas *bounding boxes* correspondentes à região do limbo de cada olho e o oclutor, conforme mostrado na Figura 23. Em seguida foi gerado um arquivo de texto para cada imagem, contendo em cada linha, um objeto de interesse que estivesse em cena. O arquivo gerado é no formato padrão de marcação YOLO.

A rede utilizou o modelo YOLOv5S pré-treinado na base microsoft COCO disponibilizado no repositório oficial de desenvolvimento da YOLOv5 ¹ e foi treinada por mais 300 épocas, visto que trata-se de um modelo pré-treinado e que quando colocado mais épocas, o ganho não era mais tão relevante. Após a realização do treinamento, o modelo foi utilizado para a detecção dos olhos e oclutor em cada *frame* dos vídeos da base.

¹ <https://github.com/ultralytics/yolov5>

Figura 23 – Esquema de marcação das imagens.



Fonte: Próprio autor.

4.2.2 Detecção do centro dos olhos

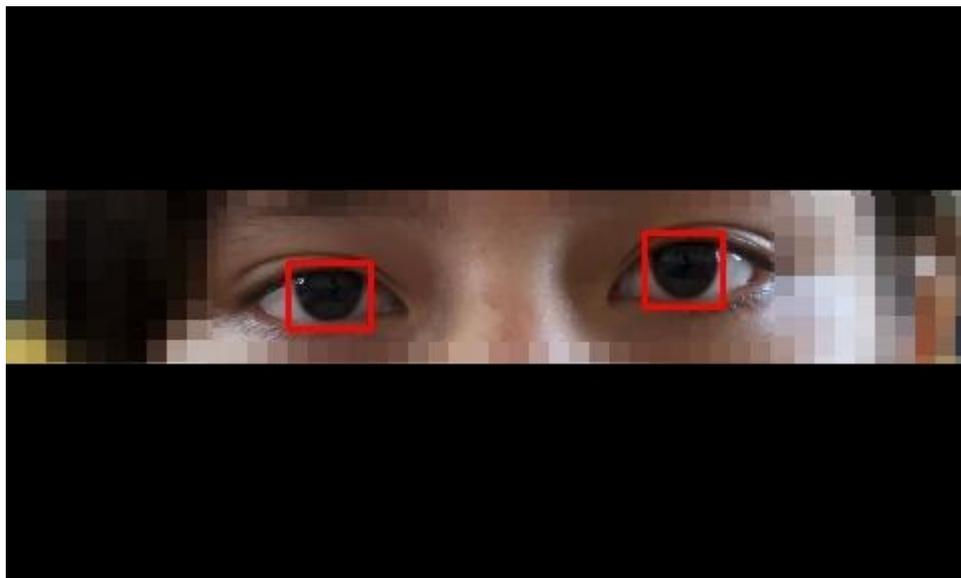
O diagnóstico realizado pelo método proposto se baseia no cálculo da diferença de posições do centro dos olhos, mais especificamente o centro do limbo, durante o decorrer do exame. Além disso como foi mostrado na Figura 23, as *bounding boxes* utilizadas nas marcações foram ajustadas para delimitar esta região. Logo, o resultado da detecção realizada a cada *frame* do vídeo retorna a delimitação da região do limbo para cada olho, esse resultado é exemplificado na Figura 24.

As detecções retornadas pela YOLOv5 são definidas pelos seguintes dados:

- *Label*, classe do objeto detectado;
- *Confidence*, confiança do objeto detectado ser de fato da classe atribuída a ele (varia entre 0 e 1);
- *Pmin*, coordenadas (X,Y) do ponto mínimo que define a *bounding box* do objeto detectado;
- *Pmax*, coordenadas (X,Y) do ponto máximo que define a *bounding box* do objeto detectado.

Aproveitando-se do resultado da detecção mostrado na Figura 24 e dos dados retornados pela rede, calcula-se as coordenadas X e Y do centro do limbo através das

Figura 24 – Detecção da região do limbo de ambos os olhos.



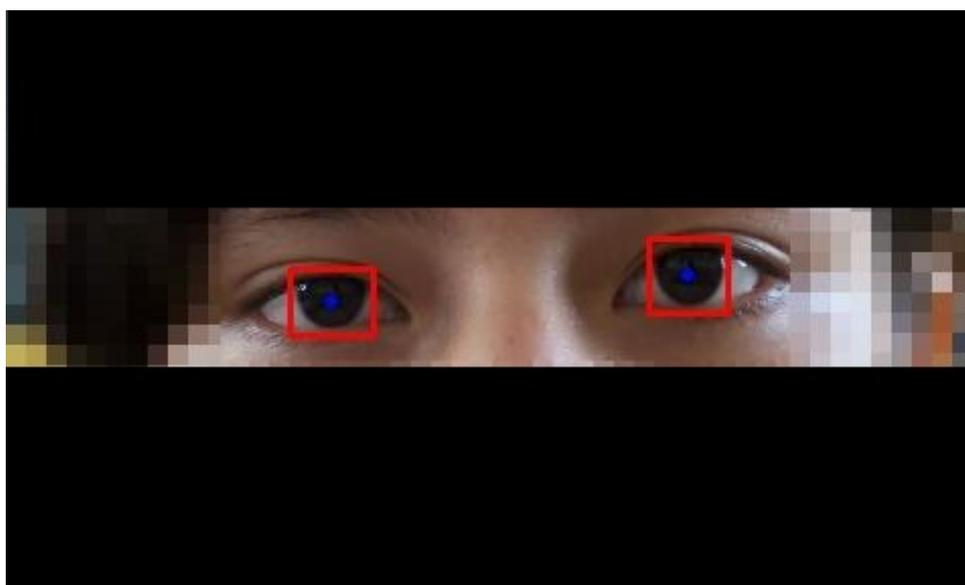
Fonte: Próprio autor.

Equações 20 e 21. O resultado obtido é apresentado na Figura 25, na qual é possível observar a representação.

$$X_{centro} = X_{min} + \frac{(X_{max} - X_{min})}{2} \quad (20)$$

$$Y_{centro} = Y_{min} + \frac{(Y_{max} - Y_{min})}{2} \quad (21)$$

Figura 25 – Definição do centro do limbo.



Fonte: Próprio autor.

4.3 Cálculo do desvio

A medida do desvio do paciente é realizada ao se comparar a posição do centro do limbo no *frame* anterior, ao ocluser ocultá-lo, e no *frame* logo em seguida, em que o ocluser deixa de ocluí-lo. Portanto, o desvio é calculado pela diferença absoluta das coordenadas do centro do olho no momento anterior a oclusão (posição inicial) e logo após a transição do ocluser para o outro olho (posição final), momento esse em que o desvio se torna aparente. Esse cálculo é realizado através das Equações 22 e 23:

$$DH = |X_{ci} - X_{cf}| \quad (22)$$

$$DV = |Y_{ci} - Y_{cf}| \quad (23)$$

onde X_{ci} e Y_{ci} são as coordenadas iniciais do centro do olho, antes do ocluser ocultá-lo, X_{cf} e Y_{cf} são as coordenadas do centro do olho após a saída do ocluser. DH e DV , são as medidas em pixels dos desvios horizontal e vertical, respectivamente.

Para a comparação futura dos resultados é preciso que o desalinhamento ocular seja transformado para dioptrias prismáticas (Δ), unidade de medida utilizada pelo especialista. A conversão é realizada através da Equação 24:

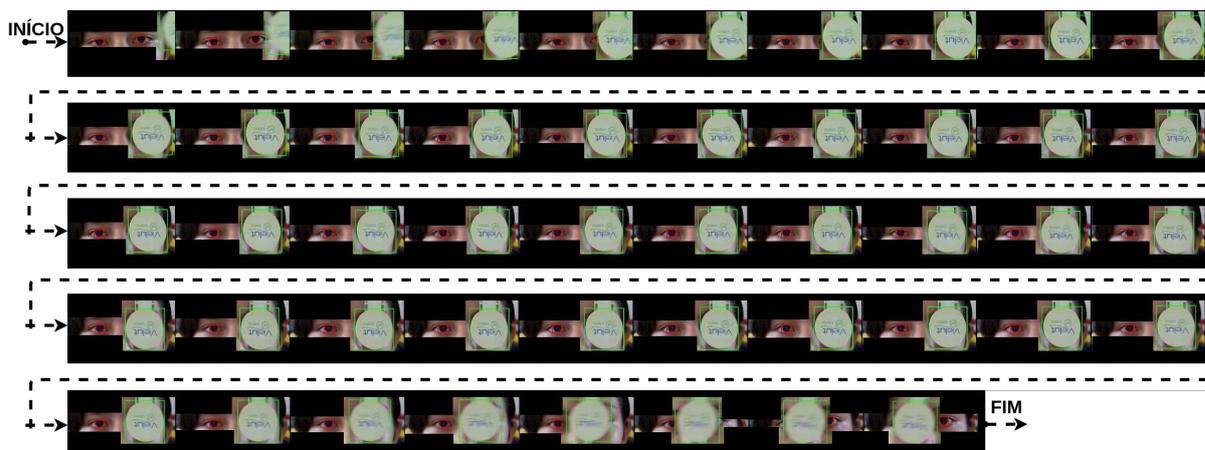
$$Dev_{\Delta} = (DE_{mm}/DE_p) * dpMM * Dev_p \quad (24)$$

onde DE_p corresponde ao valor do diâmetro do limbo do paciente detectado em *pixels*, DE_{mm} é o valor do diâmetro médio do limbo de um adulto, correspondendo a 11 mm (KHNG; OSHER, 2008), $dpMM$ trata-se de uma constante igual a 15Δ (SCHWARTZ, 2006) e representa a relação entre dioptrias por milímetros, Dev_p representa o valor do desvio calculado em *pixels* pelo método proposto e, por fim, Dev_{Δ} é o valor final em dioptrias prismáticas.

Conforme mencionado, uma das variáveis da Equação 24 é a DE_p , a qual representa o valor do diâmetro estimado do limbo do paciente. Para realizar a estimação desse valor, aproveita-se do ajuste da *bounding boxes* a região que engloba a íris, ou seja, o limbo. Para a estimação, considera-se a largura da *bounding boxes* que delimita o olho, para cada olho nos 20 primeiros *frames* do vídeo, após a passagem dos frames é calculada a média entre todas as larguras dessas *bounding boxes*.

Também é considerada a execução do exame *Cover test*, que nos vídeos que compõem a base, cada olho é ocluído diversas vezes e de forma alternada. A Figura 26 apresenta o processo de oclusão completo de um olho. No primeiro *frame* da sequência o ocluser é detectado em cena e começa a se aproximar do olho esquerdo do paciente, nos

Figura 26 – Processo de oclusão de um olho.



Fonte: Próprio autor.

frames posteriores o olho é completamente ocluído. Por fim, no último *frame* da sequência é possível observar o olho completamente desobstruído, esse processo é repetido para cada olho diversas vezes ao decorrer do vídeo.

Dessa forma, um desvio D é calculado para cada tempo T , ou seja, cada vez que um dos olhos sofre o processo de oclusão. São calculados o desvio horizontal e vertical para cada tempo T . Dada uma situação em que um olho é ocluído 5 vezes durante o vídeo, serão calculados os desvios DT_1, \dots, DT_5 . Por fim, são obtidos os seguintes valores:

- **DHE**: Média dos desvios horizontais para cada tempo (DT) para o olho esquerdo;
- **DVE**: Média dos desvios verticais para cada tempo (DT) para o olho esquerdo;
- **DHD**: Média dos desvios horizontais para cada tempo (DT) para o olho direito;
- **DVD**: Média dos desvios verticais para cada tempo (DT) para o olho direito;
- **MDH**: média dos desvios horizontais dos dois olhos;
- **MDV**: média dos desvios verticais dos dois olhos

4.4 Hardware e Software utilizados

A metodologia apresentada neste capítulo foi desenvolvida por completo na linguagem de programação *Python*, na versão 3.7 (ROSSUM; JR, 1995) em conjunto com o gerenciador de ambiente virtuais e bibliotecas *Anaconda navigator* (ANACONDA, 2020). Esta linguagem foi escolhida por oferecer grande suporte ao desenvolvimento de sistemas baseados em Aprendizado de Máquina, e ao processamento e análise de dados científicos. Dentre as principais bibliotecas em *Python* usadas no desenvolvimento desta pesquisa

destaca-se o *PyTorch* ([PASZKE et al., 2019](#)), na versão 1.7.0. O computador utilizado para o desenvolvimento de boa parte da metodologia conta com sistema operacional Windows 10, com processador Intel Core i7-7700, 16GB de memória RAM e 120 *gigabytes* de SSD. Para o treinamento da CNN YOLOv5, utilizou-se a GPU NVIDIA 1080Ti com frame buffer de 11 *gigabytes*.

5 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados obtidos nas diferentes etapas deste trabalho. A Seção 5.1 relata os resultados na etapa de detecção dos olhos e ocluser. Já a Seção 5.3 apresenta os resultados obtidos na medição dos desvios dos pacientes, comparação dos resultados com o trabalho de Valente et al. (2017), compara os resultados obtidos com o diagnóstico do especialista e por fim analisa os resultados obtidos através de métricas da literatura.

5.1 Resultados dos detecção dos olhos

Como descrito na Seção 4.2.1, a base para esse experimento contém 12.253 imagens, dessas, 70% foram utilizadas para treinamento, 20% para validação e 10% para teste. Foi utilizado um modelo pré-treinado da base *microsoft* COCO e treinado por mais 300 épocas. Esse modelo obteve 100% de acerto na detecção dos olhos, especificamente a região do limbo conforme mostrado na Figura 23 na Seção 4.2.1. Já na detecção do ocluser o modelo obteve 97% de acerto.

O modelo treinado também foi aplicado sobre os vídeos do *Dataset* PSN (Seção 4.1.3), com o objetivo de realizar uma avaliação da taxa de acertos na detecção dos olhos em uma situação mais próxima da real, ou seja, a detecção de olhos em vídeos de pacientes acometidos por patologia da visão.

O resultado da detecção realizada pelo modelo YOLOv5 na base de vídeos PSN é mostrado na Tabela 4. Os 35 vídeos da base totalizam 19966 olhos, dos quais 19488 foram detectados corretamente pela YOLOv5. Do total, apenas 478 olhos não foram detectados, fazendo com que a rede obtivesse uma assertividade de 97,61%.

Tabela 4 – Resultado modelo YOLO no *Dataset* PSN

Olhos	Acertos	Erros	Assertividade (%)
19966	19488	478	97,61

5.2 Resultados do diagnóstico para o Dataset CV-I

Como dito anteriormente, o método proposto também foi avaliado usando vídeos do trabalho de Valente et al. (2017). A comparação dos resultados do método com o diagnóstico do especialista é apresentada na Tabela 5. A tabela contém a identificação do

paciente, o diagnóstico do especialista, o desvio horizontal do método (MDH) e o desvio vertical (MDV) realizados pelo método proposto.

De acordo com o estudo do Valente et al. (2017), a base de vídeos é composta apenas por pacientes com estrabismo horizontal e pacientes categorizados como orto, ou seja, aqueles que não apresentam desvios. Quando os valores de desvio apontados pelo especialista são comparados com o desvio horizontal estimado pela técnica (MDH), o método apresenta um erro absoluto máximo de $6,96\Delta$, ainda inferior à tolerância aprovada de 8Δ . Por outro lado, ao considerar que os pacientes não apresentam desvios verticais, ou seja, um desvio igual a 0Δ , o método proposto apresenta erro absoluto vertical máximo de $5,2\Delta$, continuando abaixo da tolerância definida.

Tabela 5 – Comparação entre medidas do especialista e medição realizada pelo método proposto

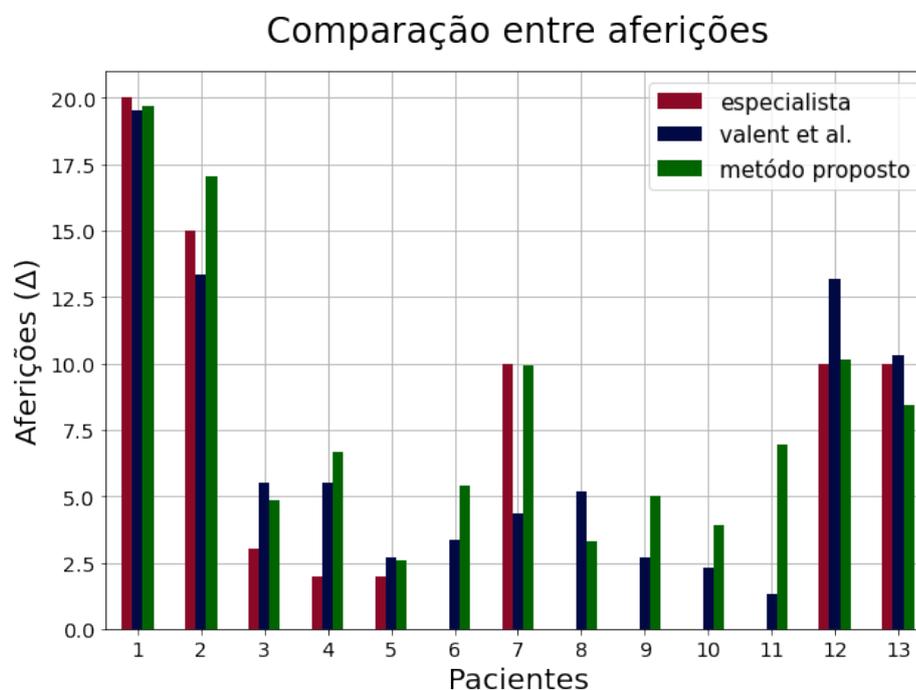
paciente	Especialista	MDH	MDV
1	20	19,69	3,235
2	15	17,04	3,47
3	3	4,87	2,71
4	2	6,65	5,20
5	2	2,59	3,03
6	orto	5,42	1,48
7	10	9,91	0,985
8	orto	3,29	1,315
9	orto	4,99	0,935
10	orto	3,89	1,95
11	orto	6,96	1,74
12	10	10,17	3,59
13	10	8,41	3,57

Ainda se considerarmos o limiar de 8Δ e utilizarmos o MDH para classificar os pacientes, o método classifica 5 pacientes como verdadeiro positivo (VP), nenhum paciente como falso negativo (FP), 8 pacientes como verdadeiro negativo (VN) e nenhum paciente como Falso Negativo (FN). Dessa forma, obtém uma acurácia igual a 100% e (95%IC = [0,77; 1]), 100% de especificidade com (95%IC = [0,66; 1]) e 100% de sensibilidade com (95%IC = [0,48; 1]) para o estrabismo horizontal, considerando uma prevalência de 2% (WILLIAMS et al., 2008). Caso seja considerado o MDH e MDV em conjunto na realização da classificação dos pacientes, para essa base de vídeos, o resultado da classificação não se alteraria.

Uma comparação entre os resultados obtidos pelo método proposto sobre o dataset CV-I com os resultados apresentados por Valente et al. (2017) e o diagnóstico do médico especialista é mostrado na Figura 27. Ao analisar o gráfico é possível observar que as medições dos dois métodos são próximas ao diagnóstico do especialista. Contudo, um caso importante acontece ao analisar o paciente de número 7, no qual o especialista aferiu um desvio de 10Δ e considerando o limiar de 8Δ o paciente deve ser classificado como doente. A metodologia de Valente et al. (2017) classifica o paciente como saudável apresentando

um erro de classificação, por outro lado, o método proposto classifica o mesmo paciente de forma correta ao aferir um desvio de $9,91\Delta$, com erro na medição de $0,09\Delta$.

Figura 27 – Comparação entre medidas de especialistas, Valente et al. (2017) e o método proposto (MDH).



Fonte: Próprio autor.

5.3 Resultados do diagnóstico para o Dataset CV-II

Com o objetivo de validar o método proposto, foram realizados experimentos utilizando a base de vídeos descrita na Seção 4.1, ao todo o método foi testado nos 57 vídeos validados e que não apresentaram problemas de aquisição. A acurácia do diagnóstico alcançado pelo método foi avaliada comparando-a com o diagnóstico apresentado pelo especialista. Em Lang (1974), é definido o conceito de esotropia de pequeno ângulo, casos em que o desvio está entre 8 e 18Δ , desvios que tem pouco impacto na rotina dos pacientes. Outro estudo foi realizado por Choi e Kushner (1998), afirma que médicos especialistas e com grau de experiência elevado apresentam um erro médio de 5 a 10Δ em suas medições. Além disso o mesmo estudo aponta que especialistas comumente aferem desvios em múltiplos de 5Δ desconsiderando em alguns casos desvios abaixo de 8Δ , nesse caso o paciente é diagnosticado como ORTO.

Na Tabela 6 foram separados os casos em que o especialista diagnosticou o paciente como estrábico. Dentre os dados que compõem a tabela estão: os identificadores dos pacientes, diagnóstico horizontal realizado pelo especialista e o valor estimado do diâmetro do limbo (DE_p). Além desses dados, também são mostrados os valores DHE , DVE , DHD ,

DVD, *MDH* e *MDV*, descritos na Seção 4.3. Ao se comparar o desvio apontado pelo especialista e compará-lo a média do desvio horizontal dos dois olhos (*MDH*), é possível observar que o método apresenta um erro máximo de $6,86\Delta$ e um erro médio absoluto de $3,75\Delta$. É possível observar que os diagnósticos realizados ficam abaixo da tolerância de 8Δ também para os pacientes que apresentam comprovadamente o estrabismo.

Tabela 6 – Comparação entre o diagnósticos de pacientes doentes.

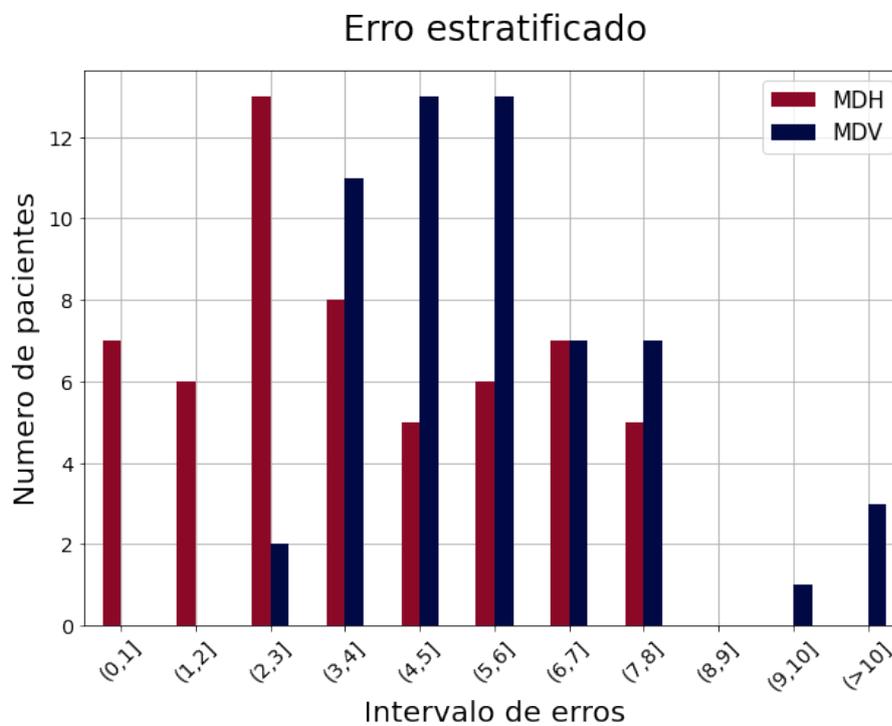
paciente	dig. especialista (horizontal)	DEp (pixels)	DHE(Δ)	DVE(Δ)	DHD(Δ)	DVD(Δ)	MDH(Δ)	MDV(Δ)
20	8,0	40,45	10,2	5,44	4,08	2,04	7,14	3,74
24	2,0	42,45	7,77	4,86	5,83	3,89	6,8	4,38
36	10,0	32	10,31	11,34	10,31	12,89	10,31	12,12
37	12,0	28,7	2,88	5,75	7,67	8,62	5,28	7,19
39	8,0	27,25	0	1,51	3,03	6,06	1,52	3,79
40	6,0	38,15	4,33	1,73	4,33	6,49	4,33	4,11
43	15,0	39,6	11,11	3,13	5,17	6,25	8,14	4,69
57	6,0	39,8	4,15	4,15	3,11	8,29	3,63	6,22

Para uma análise mais completa, a Tabela 7 apresenta os resultados obtidos pelo método proposto. A Tabela 7 é resumida graficamente na Figura 28. O gráfico mostra que o método atinge um erro de medição do desvio horizontal (*MDH*) entre 0 e 4Δ na maioria dos pacientes. Esses valores são menos da metade da tolerância de 8Δ . O método apresenta um erro máximo de $7,28\Delta$ e um Erro médio absoluto de $3,74\Delta$. Considerando a magnitude dos desvios e a tolerância de 8Δ como erro, conforme o estudo de [Choi e Kushner \(1998\)](#), a metodologia se manteve dentro da tolerância. Ao verificar os valores exibidos na Tabela 7 e observando-se os casos em que os pacientes foram classificados pelo especialista como ORTO, ou seja, um desvio abaixo de 8Δ . Logo, a metodologia obteve uma média de desvios das linhas horizontais de ambos os olhos (*MDH*) abaixo deste limiar nos 48 casos, implicando que nenhum paciente seria classificado como estrábico pelo método. Considerando o limiar de 8Δ , o método classificou 2 pacientes como verdadeiro positivo (VP), nenhum paciente como falso positivo (FP), 54 pacientes como verdadeiro negativo (VN) e 1 paciente como falso negativo (FN). Portanto, o método obteve uma sensibilidade de 66,67% com (IC 95% [0,64; 0,99]), especificidade de 100% com (IC 95% [0,93; 1]) e acurácia de 99,33% com (IC 95% [0,92; 1]) para triagem de estrabismo horizontal considerando uma prevalência de estrabismo de 2% ([WILLIAMS et al., 2008](#)).

Conforme descrito na Seção 3.6, foram realizados testes de hipóteses para a avaliar a diferença estatística entre a classificação dos pacientes pelo método proposto e as medições do especialista. Os parâmetros utilizados foram o tamanho da amostra (57 pacientes) e o nível de significância (α) de 0,05 conforme recomendado na literatura. A hipótese nula é de que não há diferença significativa entre o método proposto e o especialista. No primeiro momento foi realizado o teste sobre a classificação utilizando o *MDH* e o especialista. Portanto, utilizando o teste Z, obteve-se um valor $p = 0,2431$. Logo, a hipótese nula é aceita, afirmando que a classificação do método é semelhante estatisticamente ao

especialista.

Figura 28 – Número de pacientes por intervalo de erro para a medida horizontal (MDH) e para a medida vertical (MDV).



Fonte: Próprio autor.

Se o desvio horizontal (MDH) e vertical (MDV) são considerados em conjunto para classificar os pacientes, o método proposto classifica 2 pacientes como verdadeiro positivo (VP), 3 pacientes como (FP), 51 pacientes como (VN) e 1 paciente falso negativo (FN). Portanto, o método obteve uma sensibilidade de 66,67% com (IC 95% [0,63; 0,99]), especificidade de 94,44% com (IC 95% [0,84; 0,98]) e acurácia de 93,89% com (IC 95% [0,84; 0,98]) considerando uma prevalência de estrabismo de 2% (WILLIAMS et al., 2008). Mesmo apresentando ligeira queda nos valores das métricas, o método proposta ainda se mostra viável, principalmente para ser aplicado em triagem de pacientes. Vale ressaltar que o (MDV), desvio vertical médio do paciente 38 apresentado na Tabela 7, foi impactado pelo fato de que o paciente se moveu abruptamente a linha de visão devido a um estímulo externo ao exame, prejudicando assim os resultados obtidos. Pelo fato de a aquisição dos vídeos serem feitas em escolas, o número de pessoas saudáveis examinadas acabou sendo maior do que o número de pessoas doentes. Por outro lado, vale ressaltar que os vídeos foram adquiridos em escolas muito diferentes de um ambiente de consultório médico. Esse ambiente se aproxima muito de uma situação real de triagem, fator importante quando se objetiva a possibilidade da utilização do método em processos de triagem.

O outro teste de hipótese realizado, seguiu os mesmos parâmetros do anterior. Tamanho da amostra (57 pacientes) e o nível de significância (α) de 0,05 conforme

recomendado na literatura. A hipótese nula é de que não há diferença significativa entre o método proposto e o especialista. No entanto, agora foi utilizado o resultado da classificação quando se utiliza MDH e MDV em conjunto e o especialista. Portanto, utilizando o teste Z, obteve-se um valor $p = 1$. Logo, a hipótese nula é aceita, afirmando que a classificação do método utilizando o MDH e MDV em conjunto é semelhante estatisticamente ao especialista.

Ao se analisar os resultados obtidos, é possível constatar que o método proposto alcança o objetivo proposto na Seção 1.1, mesmo que ainda haja espaço para melhorias. Um ponto que deve ser avaliado, e caso necessário melhorado, é a estimação do diâmetro do limbo (DEp), pois, observando-se na Tabela 7 é possível notar que existe grande variação entre os valores obtidos por esta etapa do método, mesmo que não haja grande variação no protocolo de aquisição, conforme mostrado na Seção 4.3. Essa etapa é fundamental para a transformação do desvio calculado em *pixels* para dioptrias prismáticas, unidade de medida que melhor atende o especialista.

Tabela 7 – Comparação entre a avaliação do especialista e os resultados alcançados.

paciente	dig. especialista (horizontal)	DEp (pixels)	DHE(Δ)	DVE(Δ)	DHD(Δ)	DVD(Δ)	MDH(Δ)	MDV(Δ)
1	orto	44,3	2,79	2,98	1,24	5,58	2,02	4,28
2	orto	39,65	3,33	6,24	4,16	6,24	3,75	6,24
3	orto	35,2	4,69	5,86	4,69	4,69	4,69	5,28
4	orto	37,5	2,2	8,8	4,4	2,2	3,3	5,5
5	orto	38,2	8,64	6,48	4,32	4,32	6,48	5,4
6	orto	41,7	3,96	9,89	3,96	3,96	3,96	6,93
7	orto	35,35	9,89	4,67	4,67	4,67	7,28	4,67
8	orto	38,9	5,65	5,3	2,12	4,24	3,89	4,77
9	orto	47,95	6,88	9,46	6,88	3,44	6,88	6,45
10	orto	46,3	8,5	9,8	5,34	3,56	6,92	6,68
11	orto	39,85	6,21	11,04	4,14	2,07	5,18	6,56
13	orto	44,75	2,46	9,21	0	5,53	1,23	7,37
14	orto	40,65	2,71	5,07	2,03	4,06	2,37	4,57
15	orto	36,65	4,5	5,63	6,75	4,5	5,63	5,07
16	orto	41	2,68	2,01	8,04	4,02	5,36	3,02
17	orto	42,4	1,3	1,95	0	3,89	0,65	2,92
18	orto	37,9	5,8	2,17	8,71	4,35	7,26	3,26
19	orto	30,95	5,33	8	2,67	5,33	4	6,67
20	8.0	40,45	10,2	5,44	4,08	2,04	7,14	3,74
21	orto	31,75	6,5	7,28	6,93	7,79	6,72	7,54
22	orto	40,7	7,09	5,67	1,35	6,08	4,22	5,88
23	orto	42	2,95	3,14	2,62	5,89	2,79	4,52
24	2.0	42,45	7,77	4,86	5,83	3,89	6,8	4,38
25	orto	45	2,45	7,34	0	3,67	1,23	5,51
26	orto	32,9	3,35	5,02	2,51	5,02	2,93	5,02
27	orto	32,1	6,85	3,85	5,14	5,14	6	4,5
28	orto	32,3	0	5,11	5,11	5,11	2,56	5,11
29	orto	40,1	4,11	14,81	2,74	6,17	3,43	10,49
30	orto	48,55	8,36	12,92	5,67	5,1	7,01	9,01
31	orto	33,5	4,93	4,93	0	4,93	2,47	4,93
32	orto	41,45	3,98	3,98	1,99	3,98	2,99	3,98
34	orto	34,7	1,59	2,38	0	4,76	0,8	3,57
35	orto	29,6	3,72	2,79	0	5,57	1,86	4,18
36	10.0	32	10,31	11,34	10,31	12,89	10,31	12,12
37	12.0	28,7	2,88	5,75	7,67	8,62	5,28	7,19
38	orto	29,1	5,67	95,26	3,78	8,5	4,73	51,88
39	8.0	27,25	0	1,51	3,03	6,06	1,52	3,79
40	6.0	38,15	4,33	1,73	4,33	6,49	4,33	4,11
41	orto	40,4	1,02	4,9	2,72	6,12	1,87	5,51
42	orto	34	4,85	9,71	2,42	4,85	3,64	7,28
43	15.0	39,6	11,11	3,13	5,17	6,25	8,14	4,69
44	orto	38,7	0	4,26	4,26	2,13	2,13	3,2
45	orto	38,7	0	4,26	4,26	2,13	2,13	3,2
46	orto	33,1	0	6,98	0	7,47	0	7,23
47	6.0	40,1	5,14	4,11	9,6	6,17	7,37	5,14
48	orto	40,55	4,07	6,51	10,18	4,07	7,13	5,29
49	orto	41,85	2,63	0,98	7,89	3,94	5,26	2,46
50	orto	40,55	4,07	2,04	0	4,07	2,04	3,06
51	orto	38,35	5,73	3,22	8,6	4,3	7,17	3,76
52	orto	39,85	5,52	5,17	2,07	4,14	3,8	4,66
53	orto	32,37	1,27	8,16	0	7,64	0,64	7,9
54	orto	29,45	7,01	4,48	4,2	11,21	5,61	7,85
55	orto	37,2	1,78	2,96	3,33	8,87	2,56	5,92
56	orto	38,3	4,31	6,46	4,31	4,31	4,31	5,39
57	6.0	39,8	4,15	4,15	3,11	8,29	3,63	6,22
58	orto	35,9	2,76	0	1,53	6,89	2,15	3,45
59	orto	33,3	1,65	4,95	0	4,95	0,83	4,95

5.4 Discussão

O método proposto permite o diagnóstico de estrabismo em vídeos do exame Cover Test alternado. O diagnóstico é realizado através da medição do desvio do paciente e apresenta um erro baixo e no pior caso inferior ao limiar de erro de médicos especialistas (CHOI; KUSHNER, 1998).

Os métodos propostos por Almeida et al. (2015), Valente et al. (2017) e Zheng et al. (2018) necessitam de um alto controle do ambiente em que o exame é executado. Esse controle é necessário pois os métodos possuem muitos parâmetros que são impactados pela iluminação, distância do paciente para câmera e etc. Por outro lado, o método, conforme apresentado, pode ser utilizado em ambientes diversos como em consultório médico, para auxiliar os especialistas na detecção e aferição do desvio ocular, ou em ambientes escolares com a finalidade de realização de triagem.

Outros trabalhos presentes na literatura, tais como Simoes et al. (2019b) e Simoes et al. (2019a), utilizam deep learning para a detecção do estrabismo. Contudo, apenas realizam a classificação do paciente em estrábico ou saudável. No método proposto por esse trabalho, realiza-se o emprego de *Deep Learning* por ser uma solução robusta para a detecção dos olhos, o que permite não só a classificação, mas também o cálculo do desvio aproximado do paciente, permitindo um diagnóstico mais completo.

Na literatura também existem trabalhos que propõem além de softwares, hardwares para auxiliar em um melhor diagnóstico do estrabismo. Com essa abordagem, destacam-se os trabalhos de Zheng et al. (2018), que utiliza computadores, câmeras e controladores eletrônicos para a execução mais precisa do exame *cover test*. Weber et al. (2017) faz uso de um óculos com câmera infra-vermelha para identificar mais facilmente a posição dos olhos. Miao et al. (2020) e Mehringer et al. (2020) utilizam óculos de realidade virtual para emular o *cover test* alternado e alcançam bons resultados, com valores de erros muito baixos. No entanto, a utilização de hardwares complexos e caros, dificulta o acesso e uso mais amplo da metodologia. Diferente desses trabalhos, o nosso método necessita apenas de um computador para processamento dos vídeos e algum aparelho capaz de adquirir os vídeos, podendo ser até mesmo um *smartphone*.

Ainda existem trabalhos como Dericioğlu e Çerman (2019) que utilizam aplicativos de celular para a detecção do estrabismo. Essa abordagem tende a ter um baixo custo. Contudo, realizam o processamento de imagens estáticas o que possibilita a detecção apenas do estrabismo manifesto, assim como Almeida et al. (2012). Enquanto que o nosso método, utilizando vídeos do exame do Cover Test, é capaz de detectar também o estrabismo em sua forma latente.

6 CONCLUSÃO

A seguinte pesquisa destinou-se a apresentar um novo método para a detecção e diagnóstico automático do estrabismo em vídeos digitais do exame *Cover test* alternado. O método apresentado empregou uma rede neural convolucional, chamada YOLOv5. A rede foi utilizada especificamente na etapa de detecção do oclutor e da região dos olhos conhecida como limbo (Seção 4.2). Dentre os benefícios da utilização da YOLOv5 nesta etapa, destacam-se: uma etapa de detecção menos parametrizada e mais robusta quando comparada com outras abordagens da literatura (Capítulo 2); sua robustez na detecção de objetos permite a aquisição de vídeos em ambientes menos controlados, aumentando as possibilidades de aplicação do método proposto; outro benefício intrínseco a rede, é a geração de um modelo pequeno, o que abre margem para emprego do método em dispositivos com *hardwares* mais acessíveis. Para a elaboração do método, também foi realizada a aquisição e construção de uma base de vídeos do exame *Cover test* alternado. Como já mencionado, os vídeos foram adquiridos em escolas, com protocolo mais flexível, que se aproxima muito de uma situação de triagem. Dessa forma, o uso do método em triagens se torna viável, possibilitando o desafogamento e o auxílio para o médico especialista.

O método proposto foi executado sobre duas bases de vídeos do exame *cover test* (Seção 4.1) com características distintas. A primeira base de vídeos (Dataset CV-I), foi construída no trabalho de Valente et al. (2017), conta com 13 vídeos de voluntários adultos e fornecida durante o desenvolvimento deste trabalho. Esta base foi adquirida em um consultório médico, sob um protocolo mais rígido de aquisição, com controle de iluminação e outros fatores conforme explicado na Seção 4.1. A segunda base de vídeos (Dataset CV-II) foi construída durante a execução deste trabalho e é formada por vídeos de 57 crianças com autorizações dos pais ou responsáveis devidamente documentadas. Os vídeos foram adquiridos utilizando um protocolo mais flexível e em escolas públicas na cidade de São Luís - MA.

Ao se comparar os resultados obtidos pelo método proposto (Capítulo 5) com aferições de um especialista em estrabismo, o método obteve um erro máximo de $7,28\Delta$ na aferição de desvios horizontais, ficando abaixo do limiar de erro definido na literatura como 8Δ . Além disso, ao se avaliar o resultado no Dataset CV-I o método alcança uma acurácia igual a 100% e (95%IC = [0,77; 1]), 100% de especificidade com (95%IC = [0,66; 1]) e 100% de sensibilidade com (95%IC = [0,48; 1]) para o estrabismo horizontal, considerando uma prevalência de 2%. Para o Dataset CV-II o método obteve uma sensibilidade de 66,67% com (IC 95% [0,64; 0,99]), especificidade de 100% com (IC 95% [0,93; 1]) e acurácia de 99,33% com (IC 95% [0,92; 1]) quando considerado apenas a medição horizontal. Por fim,

ao se considerar as medições horizontais e verticais em conjunto o método obteve uma sensibilidade de 66,67% com (IC 95% [0,63; 0,99]), especificidade de 94,44% com (IC 95% [0,84; 0,98]) e acurácia de 93,89% com (IC 95% [0,84; 0,98]) considerando uma prevalência de estrabismo de 2%.

Desta forma, analisando o objetivo geral deste trabalho proposto na Seção 1.1, acredita-se que o mesmo foi alcançado, visto que, o método apresentado obteve resultados que condizem com ao diagnóstico do estrabismo. O mesmo vale para os objetivos específicos.

6.1 Trabalhos futuros

Apesar de ser uma área de estudo bastante explorada, fica evidente as diversas possibilidades a serem avaliadas e que podem ocasionar melhorias nos resultados obtidos. Desta forma, destacam-se alguns pontos a serem investigados em trabalhos futuros: (1) realizar experimentos em outras bases de vídeos; (2) realizar aquisição de mais vídeos de pacientes com o estrabismo; (3) utilizar a etapa de cálculo do desvio para fornecer também a classificação do estrabismo (3.2); (4) otimizar as etapas do método para que seja possível utilizá-lo em tempo real e (5) realizar a portabilidade do método, possivelmente, par um dispositivo móvel afim de facilitar seu uso em triagens. Acredita-se que as etapas 1, 2 e 3 são de suma importância para enriquecimento do trabalho. Já os itens 4 e 5 são pontos que espera-se permitir de fato o uso do método em situações reais e aumentar o seu impacto em outras áreas fora da academia.

6.2 Contribuições

A principal contribuição deste trabalho é a proposta de um método para a detecção de diagnóstico de estrabismo baseado em vídeos do exame *cover test* alternado, utilizando a rede neural convolucional YOLOv5. Outras contribuições que podem ser citadas:

- Aquisição e construção de uma base de vídeos do exame *cover test* alternado.
- Elaboração de uma revisão sistemática ampla da literatura (RSL) sobre métodos computacionais aplicados ao estrabismo.
- Criação de uma base de dados com marcações de olhos e oclisor para treinamento de redes neurais de detecção de objetos.
- Utilização de uma rede neural convolucional para a localização da região do olhos e oclisor.
- Transferência de aprendizado do modelo de CNN para detecção de olhos e oclisor

6.3 Produções científicas

A Tabela 8 apresenta o artigos científicos desenvolvidos durante o tempo desta pesquisa. Todos os trabalhos foram produzidos como autor principal, com exceção do primeiro trabalho listado, que foi como co-autor.

- T01 - Detecção de Patologias Oculares em Imagens de Reflexo Vermelho Utilizando Descritores de Cor;
- T02 - *An Automated Cover Test Approach for Strabismus Diagnosis using YOLOv5*;
- T03 - *An automated covert test approach for strabismus diagnosis using deep learning*.

Tabela 8 – Produções científicas desenvolvidas durante a pesquisa.

Status	Título	Evento	Ano	Qualis
Publicado	T01	SBCAS - XXI Simpósio Brasileiro de Computação Aplicada à Saúde	2021	B3
Aceito p/ Publicação (Retirado)	T02	IEEE - Medical Measurements & Applications	2022	A4
Submetido	T03	International Journal of Medical Informatics	2022	A1

Fonte: Próprio autor

Referências

- ADIBHATLA, V. A. et al. Applying deep learning to defect detection in printed circuit boards via a newest model of you-only-look-once. *American Institute of Mathematical Sciences (AIMS)*, 2021. Citado 2 vezes nas páginas 43 e 44.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. *2017 international conference on engineering and technology (ICET)*. [S.l.], 2017. p. 1–6. Citado na página 37.
- ALMEIDA, J. D. S. D. et al. Computer-aided methodology for syndromic strabismus diagnosis. *Journal of digital imaging*, Springer, v. 28, n. 4, p. 462–473, 2015. Citado 4 vezes nas páginas 21, 24, 27 e 68.
- ALMEIDA, J. D. S. de et al. Computational methodology for automatic detection of strabismus in digital images through hirschberg test. *Computers in biology and medicine*, Elsevier, v. 42, n. 1, p. 135–146, 2012. Citado 4 vezes nas páginas 21, 22, 27 e 68.
- ANACONDA. *Anaconda Software Distribution*. [S.l.]: Anaconda Inc., 2020. Disponível: <<https://docs.anaconda.com/>>. (Acessado em 28/05/2022). Citado na página 59.
- BAIÃO, A. F. *Simulador óptico dinâmico do olho humano*. Tese (Doutorado) — Faculdade de Ciências e Tecnologia, 2013. Citado 2 vezes nas páginas 28 e 29.
- BOARETTO, A. E. A evolução da população mundial, da oferta de alimentos e das ciências agrárias. *Revista Ceres*, Universidade Federal de Viçosa, v. 56, n. 4, p. 513–526, 2009. Citado na página 15.
- BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv, 2020. Disponível em: <<https://arxiv.org/abs/2004.10934>>. Citado 2 vezes nas páginas 42 e 43.
- BOMMIREDDY, T.; TAYLOR, K.; CLARKE, M. P. Assessing strabismus in children. *Paediatrics and Child Health*, v. 30, n. 1, p. 14–18, 2020. ISSN 1751-7222. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S175172221930232X>>. Citado na página 31.
- BOURNE, R. et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study. *The Lancet global health*, Elsevier, v. 9, n. 2, p. e130–e143, 2021. Citado na página 15.
- CBO, C. b. d. o. *Censo 2021*. 2021. <https://cbo.net.br/2020/admin/docs_upload/034327Censo2021.pdf>. (Accessed on 05/12/2022). Citado na página 16.
- CHOI, R. Y.; KUSHNER, B. J. The accuracy of experienced strabismologists using the hirschberg and krimsky tests. *Ophthalmology*, Elsevier, v. 105, n. 7, p. 1301–1306, 1998. Citado 3 vezes nas páginas 63, 64 e 68.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 40.

- COSTA, I. *Estrabismos latentes encontrados em crianças dos 0 aos 12 anos de idade, observados no serviço de oftalmologia do hospital baptista de souza nos últimos 5 anos*. Dissertação (B.S. thesis) — Universidade do Mindelo, 2017. Citado na página 30.
- COSTA, P. et al. Identificando a paralisia do sexto nervo Ótico em videos através de redes neurais convolucionais e curvatura isophotes. In: *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2019. p. 246–257. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/6258>>. Citado 2 vezes nas páginas 52 e 54.
- CRUZ, L. B. da et al. Kidney tumor segmentation from computed tomography images using deeplabv3+ 2.5 d model. *Expert Systems with Applications*, Elsevier, v. 192, p. 116270, 2022. Citado na página 34.
- DERICIOĞLU, V.; ÇERMAN, E. Quantitative measurement of horizontal strabismus with digital photography. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, Elsevier, v. 23, n. 1, p. 18–e1, 2019. Citado 5 vezes nas páginas 21, 25, 27, 54 e 68.
- DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. *Computer Science Review*, Elsevier, v. 40, p. 100379, 2021. Citado na página 34.
- FARUK, H. O. *Strabismus Simulator - American Academy of Ophthalmology*. 2015. <<https://www.aaopt.org/interactive-tool/strabismus-simulator>>. (Acessado em 22/03/2022). Citado 2 vezes nas páginas 32 e 33.
- GARRITY, J. *Estrutura e função dos olhos - Distúrbios oftalmológicos - Manual MSD Versão Saúde para a Família*. 2019. <<https://www.msmanuals.com/pt-br/casa/dist%C3%BArbios-oftalmol%C3%B3gicos/biologia-dos-olhos/estrutura-e-fun%C3%A7%C3%A3o-dos-olhos>>. (Acessado em 21/03/2022). Citado na página 31.
- GUO, Y. et al. Deep learning for visual understanding: A review. *Neurocomputing*, Elsevier, v. 187, p. 27–48, 2016. Citado 3 vezes nas páginas 37, 38 e 39.
- HATT, S. R. et al. The effects of strabismus on quality of life in adults. *American Journal of Ophthalmology*, v. 144, n. 5, p. 643–647, 2007. ISSN 0002-9394. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0002939407006101>>. Citado na página 31.
- HEALTHENGINE. *Human eye anatomy and how Vision Works Information: Myvmc*. 2007. <<https://healthengine.com.au/info/the-eye-and-vision>>. (Acessado em 28/05/2022). Citado na página 28.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks*, Elsevier, v. 3, n. 5, p. 551–560, 1990. Citado na página 35.
- JOCHER, G. et al. *ultralytics/yolov5: v4.0 - nn.SiLU() activations, Weights & Biases logging, PyTorch Hub integration*. Zenodo, 2021. Disponível em: <<https://doi.org/10.5281/zenodo.4418161>>. Citado 2 vezes nas páginas 42 e 44.
- KHNG, C.; OSHER, R. H. Evaluation of the relationship between corneal diameter and lens diameter. *Journal of Cataract & Refractive Surgery*, Elsevier, v. 34, n. 3, p. 475–479, 2008. Citado na página 58.

- KOVÁCS, Z. L. *Redes neurais artificiais*. [S.l.]: Editora Livraria da Física, 2002. Citado na página 34.
- LANG, J. Management of microtropia. *The British journal of ophthalmology*, BMJ Publishing Group, v. 58, n. 3, p. 281, 1974. Citado na página 63.
- LEHMANN, E. L.; ROMANO, J. P. *Testing statistical hypotheses*. [S.l.]: Springer Nature, 2022. Citado na página 50.
- LEITMAN, M. W. *Manual for eye examination and diagnosis*. [S.l.]: John Wiley & Sons, 2021. Citado na página 30.
- LIMA, A. A. et al. Mask overlaying: a deep learning approach for individual optic cup segmentation from fundus image. In: IEEE. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.], 2020. p. 99–104. Citado na página 34.
- LU, J. et al. Automated strabismus detection for telemedicine applications. *arXiv preprint arXiv:1809.02940*, 2018. Citado 3 vezes nas páginas 21, 23 e 27.
- MACIEL, V. D.; MATIAS, B. C.; SARINHO, V. T. Uma revisao sistemática da literatura de jogos digitais voltados para o tdah. In: SBC. *Anais do XIX Simposio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2019. p. 130–140. Citado na página 19.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 34.
- MEHRINGER, W. A. et al. An image-based method for measuring strabismus in virtual reality. In: IEEE. *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. [S.l.], 2020. p. 5–12. Citado 5 vezes nas páginas 21, 26, 27, 54 e 68.
- MIAO, Y. et al. Virtual reality-based measurement of ocular deviation in strabismus. *Computer methods and programs in biomedicine*, Elsevier, v. 185, p. 105132, 2020. Citado 5 vezes nas páginas 21, 26, 27, 54 e 68.
- ORGANIZATION, W. H. et al. *Global initiative for the elimination of avoidable blindness*. [S.l.], 2000. Citado na página 15.
- PASZKE, A. et al. Pytorch: An imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019. p. 8024–8035. Disponível em: <<http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>>. Citado na página 60.
- PEDRINI, H.; SCHWARTZ, W. R. *Análise de imagens digitais: princípios, algoritmos e aplicações*. [S.l.]: Cengage Learning, 2008. Citado na página 38.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788. Citado 3 vezes nas páginas 40, 41 e 42.

- REDMON, J.; FARHADI, A. Yolo9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 7263–7271. Citado na página 42.
- REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. Citado na página 42.
- REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, v. 28, 2015. Citado na página 40.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 34.
- ROSSUM, G. V.; JR, F. L. D. *Python reference manual*. [S.l.]: Centrum voor Wiskunde en Informatica Amsterdam, 1995. Citado na página 59.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Citado na página 36.
- SCHWARTZ, G. S. *The eye exam: a complete guide*. [S.l.]: Slack Incorporated, 2006. Citado na página 58.
- SEO, M. W. et al. The automated diagnosis of strabismus using an infrared camera. In: SPRINGER. *6th European Conference of the International Federation for Medical and Biological Engineering*. [S.l.], 2015. p. 142–145. Citado 3 vezes nas páginas 21, 22 e 27.
- SIMOES, T. d. O. et al. Automatic ocular alignment evaluation for strabismus detection using index of phylogenetic diversity and support vector machine. In: SBC. *Anais do XV Workshop de Visão Computacional*. [S.l.], 2019. p. 73–78. Citado 4 vezes nas páginas 21, 25, 27 e 68.
- SIMOES, T. de O. et al. Automatic ocular alignment evaluation for strabismus detection using u-net and resnet networks. In: IEEE. *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.], 2019. p. 239–244. Citado 4 vezes nas páginas 21, 25, 27 e 68.
- TRA, V. et al. Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal levenberg-marquardt algorithm. *Sensors*, v. 17, p. 2834, 12 2017. Citado na página 38.
- TRABOULSI, E. I.; UTZ, V. M. *Practical Management of Pediatric Ocular Disorders and Strabismus*. [S.l.]: Springer, 2016. Citado na página 32.
- TURNER, m. *The sclera and its covering – Anatomy of the Anterior Eye for Ocularists*. 2012. <<https://eyeanatomyblog.wordpress.com/2012/10/21/the-sclera-and-its-coverings/>>. (Acessado em 21/03/2022). Citado na página 29.
- UIJLINGS, J. R. et al. Selective search for object recognition. *International journal of computer vision*, Springer, v. 104, n. 2, p. 154–171, 2013. Citado na página 40.

- VALENTE, T. L. A. et al. Automatic diagnosis of strabismus in digital videos through cover test. *Computer methods and programs in biomedicine*, Elsevier, v. 140, p. 295–305, 2017. Citado 13 vezes nas páginas 10, 21, 24, 26, 27, 52, 53, 54, 61, 62, 63, 68 e 69.
- WANG, C.-Y. et al. Cspnet: A new backbone that can enhance learning capability of cnn. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. [S.l.: s.n.], 2020. p. 390–391. Citado na página 45.
- WEBER, K. P. et al. Strabismus measurements with novel video goggles. *Ophthalmology*, Elsevier, v. 124, n. 12, p. 1849–1856, 2017. Citado 6 vezes nas páginas 21, 24, 26, 27, 54 e 68.
- WHO. World report on vision. World Health Organization, 2019. Citado na página 15.
- WILLIAMS, C. et al. Prevalence and risk factors for common vision problems in children: data from the alspac study. *British Journal of Ophthalmology*, BMJ Publishing Group Ltd, v. 92, n. 7, p. 959–964, 2008. Citado 3 vezes nas páginas 62, 64 e 65.
- ZHENG, C. et al. Detection of referable horizontal strabismus in children’s primary gaze photographs using deep learning. *Translational vision science & technology*, The Association for Research in Vision and Ophthalmology, v. 10, n. 1, p. 33–33, 2021. Citado na página 21.
- ZHENG, Y. et al. An automatic stimulus and synchronous tracking system for strabismus assessment based on cover test. In: IEEE. *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. [S.l.], 2018. v. 3, p. 123–127. Citado 7 vezes nas páginas 21, 23, 25, 26, 27, 54 e 68.
- ZOLKIFLI, N. S. et al. Range detection of strabismus based on the distance and coordinates of the iris. *International Journal of Advanced Computer Systems and Software Engineering*, v. 2, n. 1, p. 23–29, 2021. Citado 3 vezes nas páginas 21, 23 e 27.
- ÁVILA, M.; ALVES, M. R.; NISHI, M. *As condições de saúde ocular no brasil 2015*. 1ª ed. [S.l.]: Conselho brasileiro de oftalmologia, 2015. Citado na página 15.